Theoretical Problems in Image Analysis

Ingeborg Koch

May 1991

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University

The work described in Chapters 2 to 4 is my own, except where otherwise indicated. The research was carried out under the supervision and guidance of Professor Peter Hall, and papers resulting from this work have been published jointly with him.

Ingeborg Koch

Acknowledgements

It gives me great pleasure to thank Professor Peter Hall for being an excellent supervisor. I greatly benefitted from his insight, knowledge and enthusiasm in many stimulating discussions.

I would also like to thank my husband, Alun Pope, for his encouragement and patience during trying times.

For Alun, Graeme and Reiner

Abstract

This thesis consists of four chapters, which deal with the estimation of images from noisy data. Chapter 1 contains a review of known and related results.

Chapter 2 deals with continuous image models in which the the true images have been systematically degraded by a known point spread function and stochastically degraded by a second order stationary random field. Partial Fourier inversion is shown to be an optimal image processing method when the noise has short-range dependence. Our main theorem shows how to construct optimal smoothing sets for the Fourier inversion and gives an optimal rate of convergence uniformly over classes of images. For specific classes of point spread functions and true images, explicit formulae for the optimal choice of the smoothing set and the associated rate of convergence have been derived in terms of the smoothness of the true image and the point spread function.

In Chapter 3, extensions of the method of cross-validation for selection of the smoothing parameter are considered in connection with discretely defined blurred and noisy images. Two generalisations of cross-validation are considered: the first ignores the blur, while the second takes careful account of the blur. Both methods are shown to result in asymptotically optimal performance provided the amount of blur does not exceed a certain level. For a specific class of point spread functions and smooth images, precise bounds for the admissible amount of blur as well as the dependence of the performance on sample size, point spread function and image smoothness are given. The first method is shown to be superior to the second, as measured by mean square error. For the first method we also show how the mean square error results can be extended to almost sure results.

The last chapter examines smoothing by local medians for images which are degraded by random noise. The mean square error of the median smoother is calculated and it is shown that the median smoother performs asymptotically as well as the local mean. The optimal window size of the median smoother is given in terms of the sample size and the dimension of the image. The rate of convergence decreases as the dimension increases, and its functional dependence on the dimension changes when the dimension is 4 or larger.

Contents

1 Introduction **5**

1.1 Background and Motivation 5

1.2 Notational Preliminaries 9

1.3 Fourier Inversion Techniques 11

1.3.1 Definitions and Properties of Fourier Transforms 11

1.3.2 Fourier Deconvolution 13

1.4 Cross-Validation in Nonparametric Regression 14

1.4.1 Performance Measures and Rates of Convergence 14

1.4.2 Choice of Smoothing Method 16

1.4.3 Selection of the Smoothing Parameter by Cross-Validation 17

1.5 M-Estimation 18

2 Partial Fourier Inversion **20**

2.1 Introduction 20

2.2 Models for Images and Image Degradation 21

2.2.1 Image Models 22

2.2.2 Models for the Blur 23

2.3 Models for Correlated Noise 26

2.4 Main Results 30

2.4.1 Bounds for the Mean Square Error 30
2.4.2 Optimality of the Partial Fourier Inversion Estimator 33
2.4.3 Discussion of Results 35
2.5 Applications 36
2.5.1 Rapidly Decreasing Fourier Kernels 36
2.5.2 Fourier Kernels with Zeroes 41
2.6 Proofs 43
2.6.1 Proof of Proposition 2.2 43
2.6.2 Proof of Proposition 2.4 44
2.6.3 Proof of Theorem 2.6 48
2.6.4 Proof of Theorem 2.7 52

3 Cross-Validation **59**

3.1 Introduction 59
3.2 Mathematical Models and Cross-Validation in Image Analysis 61
3.2.1 Mathematical Models for the Observations 61
3.2.2 The Estimator $\hat{t}$ 63
3.2.3 Two Approaches to Cross-Validation 64
3.3 Results 67
3.3.1 Properties of MSSE and SSE 68
3.3.2 Mean Square Results for the Cross-Validation Methods 71
3.3.3 Discussion of Results 73
3.3.4 Uniform *a.s.* Results for the Naive CV Method 74
3.3.5 Numerical Examples 77
3.4 Proofs 81

3.4.1 Some Technical Lemmas 81

3.4.2 Proof of Proposition 3.1 89

3.4.3 Proof of Proposition 3.4 91

3.4.4 Proof of Theorem 3.5 95

3.4.5 Proof of Theorem 3.6 101

3.4.6 Proof of Proposition 3.7 108

3.4.7 Proof of Theorem 3.8 117

3.4.8 Proof of Proposition 3.9 127

3.5 Appendix: Rosenthal's Inequality 130

4 Median Smoothing **133**

4.1 Introduction 133

4.2 Image Models and the Median Smoother 135

4.2.1 Models for the Observations 135

4.2.2 The Median Smoother $\widehat{T}$ 137

4.2.3 Approximations to the Median Smoother 138

4.3 Results 140

4.3.1 Convergence of $\widehat{T}^{\nu}$ to $\widehat{T}$ 140

4.3.2 The Mean Square Error of the $\widehat{T}^{\nu}$ 141

4.3.3 The Mean Square Error of $\widehat{T}$ 143

4.3.4 Discussion of Results 144

4.4 Proofs 146

4.4.1 Proof of Proposition 4.1 146

4.4.2 Proof of Proposition 4.2 148

4.4.3 Proof of Proposition 4.3 156

4.4.4 Proof of Theorem 4.4 162

References **165**

Chapter 1

Introduction

This thesis deals with theoretical aspects of some statistical problems which arise in the analysis and processing of images.

In Chapters 2 and 3 we look at blurred and noisy images and consider asymptotically optimal ways of deblurring the observed data. This is followed by proposals of criteria for determining the amount of deblurring needed. The method used for deblurring is partial Fourier inversion and the amount of deblurring required is determined by extensions of the method of cross-validation.

The last chapter explores the non-linear method of local medians as an estimation technique for noise-degraded images, and describes how the window size for the local median must be chosen to provide asymptotically optimal performance.

In these three chapters we present our asymptotic results in the form of rates of convergence of the estimator to the true image with respect to an appropriate distance measure.

1.1 Background and Motivation

Images come in many different shapes and sizes and arise in a large number of areas including science, engineering and medicine. The interpretation of an observed image is not always apparent from the observed data, since the true image has often been degraded in the imaging process, and may therefore only be available in a disguised form.

It is the task of image processing methods to uncover the true object and to obtain desired information about it from the observed image. Since the mid-fifties, when the processing of pictorial information by computer really began, many different methods of analysing and processing images have been developed with the aim of making the observed image more accessible to the human observer.

Many of the common processing techniques such as digitisation, coding, data compression, image enhancement and restoration, and image segmentation are described in Rosenfeld and Kak (1976). An account of image reconstruction methods, including tomographic techniques, can be found in Chapter 5 of Hall (1979). These methods have been enjoying great popularity in particular in medical imaging. A statistical approach which regards the true images as Markov random fields can be found in Besag (1986).

From the large number of areas within image processing, we are interested in image enhancement and in particular in image restoration. Techniques used in these areas include 'cosmetic' operations for image enhancement to obtain a subjectively more pleasing image, as well as objective and automatic estimation procedures, whose goal is determination of the true image. Image restoration encompasses deblurring and smoothing as well as noise removal, and is applicable to many different types of observed data.

Before we focus our attention on some specific methods, it will be expedient to look at the types of images we shall be concerned with in this thesis. Let T denote the true image and let ϵ denote random noise. Put

$$Y_j = T_j + \epsilon_j, \tag{1.1.1}$$

for j belonging to some index set in $\mathbb{Z}^d$, for $d \geq 1$. In this case the true image has been distorted only by the addition of random noise. Examples of observed images of this type are picture transmission or the scanning of images by television cameras (see Section 6.1.4 of Rosenfeld and Kak (1976)).

In many applications, the true image cannot be seen directly, as in (1.1.1), but is transformed in the imaging process, resulting in indirect observations X_j given by

$$X_j = f(T)_j,$$

for j in some index set. This transformation f may be quite arbitrary, and need not be linear or deterministic. However, in applications it is often reasonable to assume that f can be approximately described by a combination of a linear operator acting on T and additive noise which is independent of T. We restrict attention to observations of this kind and assume that the linear operator is a convolution operator K. Our observations are then of the form

$$X_j = K(T)_j + \epsilon_j = (k \star T)_j + \epsilon_j, \tag{1.1.2}$$

where ϵ denotes random noise, $\star$ denotes convolution, and k denotes the convolution kernel associated with K. In contrast to the observations Y_j given by (1.1.1), we shall sometimes refer to the observations X_j of (1.1.2) as *indirect* observations and to $k \star T$ as the *blurred* image. One may conveniently think of K as modelling the effect of the apparatus used in the imaging process. We shall assume that K (or, equivalently, k) is known and deterministic. These assumptions on K do not conflict with reality as the apparatus effect is often known or can be determined accurately. The function k may be regarded as a measure of degradation of the true image, since it describes the blurring or spreading out of the true image. In the engineering literature, k is often called the *point spread function* or the *impulse response* (see Section 2.1.1 of Rosenfeld and Kak (1976)).

For images as in (1.1.2), engineers use a wide variety of restoration methods, including inverse filtering, least squares filtering and constrained deconvolution, to name but a few. For descriptions of these and other deblurring methods, see Chapters 6 and 7 of Rosenfeld and Kak (1976). A more mathematical framework for inverse filtering can be found in Twomey (1965). These restoration methods explicitly exploit the fact that the convolution kernel is known (or can be measured accurately) and can therefore—at least in theory—be removed. Conceptually the simplest method is that of inverse filtering. It also provides a basis for some of the other methods, which address the shortcomings and the arbitrariness of inverse filtering in different ways. In Section 1.3 we describe the technique of inverse filtering, also known as Fourier deconvolution, and point out how it can fail. We then introduce the concepts of smoothing parameter and partial Fourier inversion and show how they can be used as a way of overcoming the instabilities associated with Fourier deconvolution. The choice of the smoothing parameter is intimately connected with the domain of definition of the Fourier deconvolution. This choice is crucial as it determines the compromise between the smoothness of the estimate and its faithfulness to the true image. In many practical applications of Fourier deconvolution, the smoothing parameter is chosen in an ad hoc manner or by exploiting specific properties of the underlying physical model. However, because of the trade-off between smoothness and faithfulness of the estimate, or between bias and variance, it is important to have a rigorous and objective basis for the selection of the smoothing parameter. We use a specific distance measure as the basis for the selection criterion for the smoothing parameter and then establish the optimality of partial Fourier inversion as a technique for estimating T. This is the topic discussed in Chapter 2. Our results are of an asymptotic nature; we treat a large class of realistic point spread functions simultaneously and allow for correlation in the error, in this way extending previous work in this area (see Hall (1990)).

From the point of view of numerical analysis, the blurred image $k \star T$ of (1.1.2) is regarded as a Fredholm integral of the first kind, and the problem of recovering T from $k \star T$ is usually ill-posed in the sense of Hadamard (see Bertero (1986)), that is, existence, uniqueness or continuous dependence of the solution on the data fails to hold. The method of regularisation (see Lukas (1980)) has been proposed as a solution to ill-posed problems. It extends an earlier method suggested by Tikhonov (1963a,1963b) (see also Tikhonov and Arsenin (1977)): A regularised solution $\widehat{T}_\alpha$, $\alpha > 0$, is the minimiser over a class of functions $\tilde{T}$ of

$$\|X - k \star \tilde{T}\|^2 + \alpha\Lambda(\tilde{T}), \tag{1.1.3}$$

where Λ denotes a regulariser, such as the linearised curvature functional or an entropy functional. (For more details on regularisers see Titterington (1985) and Koch and Anderssen (1986).) Although regularisation yields a unique solution, from a statistical point of view this is not enough, since

- the blurred image $k \star T$ is usually also degraded by the addition of noise as in (1.1.2); and
- the question of how to choose the regularisation or smoothing parameter α is not addressed.

For images as in (1.1.1), cross-validation provides an automatic and objective way of selecting the smoothing parameter. A brief account of cross-validation in nonparametric regression and references to the literature are given in Section 1.4. For indirect images of the form (1.1.2) no analogous method for choosing the smoothing parameter exists as yet, although the problem has been addressed by some authors (see Thompson *et al.* (1990) and Koch and Tarlowski (1986)). In Chapter 3 of the thesis, we propose generalisations of cross-validation to indirect images X_j which exploit the fact that partial Fourier inversion yields optimal estimates under suitable assumptions. The presence of the blur in the image adds another degree of complexity to the problem of giving a rigorous treatment of cross-validation. To make our analysis mathematically more tractable, we therefore restrict the blur to a specific class of point spread functions. This enables us to give precise bounds on the admissible amount of blur in the observed image. Under these conditions on the blur we show that cross-validation performs asymptotically optimally, and cross-validation fails if the blur exceeds these bounds.

So far we have been concerned with linear methods for estimating the true image. In some areas of image processing, for example, when one is interested in locating and preserving edges or in reducing the impulse noise and periodic interference patterns, non-linear techniques, and in particular, median filtering have been found to perform better than linear methods. Tukey (1977) was among the first to suggest the use of running medians as a way of smoothing signals. Since then, the running median or median filter has gained in popularity, and in some areas has even become an effective alternative to linear smoothing techniques (see Bovik *et al.* (1983)). Fast algorithms for one- and two-dimensional median filtering have been developed (see Huang *et al.* (1979) and references therein). These developments and the interest in median filtering in the engineering literature are complemented by a renewed interest in least absolute deviation fitting in optimisation and numerical analysis. A further impact on the advance of methods involving medians came from robust statistics; the median smoother is a far more robust estimator than the mean, for example, in the presence of outlier contamination (see Hampel *et al.* (1986)).

The different areas of research in which medians arise—in the form of running medians or median filters—have developed quite independently so far and little cross-fertilisation exists. In a statistical context, least absolute deviation estimators have been employed in linear regression and time series, for example, and comparison with the mean and some Huber M-estimators exist for various regression models (see Chapter 2 of Bloomfield and Steiger (1983)). In the engineering sciences, on the other hand, median filtering is still very much a practical tool, and despite growing popularity and range of applications, theoretical results are still scarce, and when they exist, are usually of a special rather than a general nature. This may be due to the sentiment expressed by Huang (1981) that a 'theoretical analysis of median filters is very difficult', which precedes papers by Tyan on deterministic properties of median filters and by Justusson on properties of median filters applied to pure noise.

The great appeal of median filters in the engineering sciences combined with the lack of a theoretical foundation has motivated us to look at median filters or median smoothers from a more theoretical and statistical point of view in Chapter 4. Unlike Justusson (see Huang (1981)), who considers pure noise, we are interested in non-constant images which are degraded by noise and consider observations of the form (1.1.1). For such data we derive an expression for the asymptotic rate of convergence of mean square error of the median smoother as the sample size increases. From this expression we deduce the asymptotically optimal window size and the corresponding rate of convergence. To prove our results we use the framework of M-estimators and robust statistics, which is briefly described in Section 1.5. Our analysis also shows how the rate of convergence depends on the dimension d of the observations, and it may be rather unexpected to observe that the rate depends on d in different functional forms depending on whether d is less than or greater than four.

1.2 Notational Preliminaries

In this section we summarize notation that may not be standard.

Norms and the inner product. For $x = (x_1, \ldots, x_d)^T \in \mathbf{R}^d$, we use the following norms:

$$|x| = \sum_{i=1}^{d} |x_i|$$

$$\|x\| = \left\{ \sum_{i=1}^{d} |x_i|^2 \right\}^{1/2}$$

$$\|x\|_\infty \;=\; \sup_i |x_i|.$$

Sometimes we shall also write $\|x\|_1$ for $|x|$ and $\|x\|_2$ for $\|x\|$. For vectors x, $y \in \mathbf{R}^d$, the inner product of x and y is

$$\langle x, y\rangle = \sum_{i=1}^{d} x_i y_i.$$

Spaces of functions. Our images and point spread functions may belong to the following spaces:

$$\begin{aligned}
L^p(\mathbf{R}^d) &= \{f : \mathbf{R}^d \to \mathbb{C} \,:\, \int |f|^p < \infty\} \quad \text{p=1,2}\\
L^p(\mathbf{R}^d, \mathbf{R}) &= \{f : \mathbf{R}^d \to \mathbf{R} \,:\, \int |f|^p < \infty\} \quad \text{p=1,2}\\
\ell^p(\mathbb{Z}^d) &= \{f : \mathbb{Z}^d \to \mathbb{C} \,:\, \sum_{k\in\mathbb{Z}^d} |f(k)|^p < \infty\} \quad \text{p=1,2}\\
C^a(\mathbf{R}^d) &= \{f : \mathbf{R}^d \to \mathbb{C} \,:\, f \text{ is } a \text{ times continuously differentiable}\}\\
C^\infty(\mathbf{R}^d) &= \{f : \mathbf{R}^d \to \mathbb{C} \,:\, f \text{ is infinitely often differentiable}\}.
\end{aligned}$$

Functional relationships. Let f and g denote positive real-valued functions defined on $\mathbf{R}^d$. We write

$$f \asymp g$$

if there exists constants c_1 and $c_2 > 0$ such that $c_1 g(x) \le f(x) \le c_2 g(x)$ for all $x \in \mathbf{R}^d$; and

$$f(x) \asymp g(x) \quad \text{as } x \to x_0$$

if $f(x)/g(x)$ and $g(x)/f(x)$ are bounded as $x \to x_0$.

Let a_n and b_n, $n = 1, 2, \ldots$, denote sequences of positive real numbers. We write

$$a_n = O(b_n)$$

if there exists a constant $c > 0$ such that $a_n/b_n \le c$ as $n \to \infty$. We write

$$a_n = o(b_n)$$

if $a_n/b_n \to 0$ as $n \to \infty$; and

$$a_n \sim b_n$$

if $a_n/b_n \to 1$ as $n \to \infty$.

For a set A, $\mathcal{I}_A$ denotes the indicator function of A.

1.3 Fourier Inversion Techniques

Fourier transforms apply to both discretely and continuously defined functions, and we consider both cases simultaneously in this section.

1.3.1 Definitions and Properties of Fourier Transforms

Let X denote a blurred image B degraded by additive random noise ϵ, so

$$X = B + \epsilon. \tag{1.3.1}$$

Here X, B, ϵ and T (below) either all denote functions defined on $\mathbf{R}^d$ or all denote functions defined on $\mathbb{Z}^d$. We assume that

$$B = H \star T, \tag{1.3.2}$$

where T denotes the true image, H denotes the blur or point spread function and $\star$ denotes the convolution product, which is defined in the following way. For F, $G \in L^1(\mathbf{R}^d)$,

$$(F \star G)(x) = \int_{\mathbf{R}^d} F(y)G(x-y)\, dy \quad \text{for } x \in \mathbf{R}^d; \tag{1.3.3}$$

and for f, $g \in \ell^1(\mathbb{Z}^d)$,

$$(f \star g)(j) = \sum_{k \in \mathbb{Z}^d} f(k)g(k-j) \quad \text{for } j \in \mathbb{Z}^d. \tag{1.3.4}$$

For indirect observations of the form (1.3.2) with known function H, Fourier deconvolution represents a natural way of recovering T from B. The method described in Subsection 1.3.2 works for discretely as well as continuously defined functions; the choice of the appropriate Fourier transform is the main distinguishing feature. For $F \in L^1(\mathbf{R}^d)$, the *d-dimensional Fourier transform* ϕ of F is given for $\theta \in \mathbf{R}^d$ by

$$\phi(\theta) = \int_{\mathbf{R}^d} F(x)e^{i\langle x,\theta\rangle}\, dx, \tag{1.3.5}$$

and if $\phi \in L^1(\mathbf{R}^d)$, the *inverse Fourier transform* $\check{\phi}$ of ϕ is defined for $x \in \mathbf{R}^d$ by

$$\check{\phi}(x) = (2\pi)^{-d} \int_{\mathbf{R}^d} \phi(\theta)e^{-i\langle x,\theta\rangle}\, d\theta. \tag{1.3.6}$$

Sometimes it will be convenient to write $\phi = FT(F)$ and $\check{\phi} = IFT(\phi)$.

Similarly, for $f \in l^1(\mathbb{Z}^d)$, the *discrete d-dimensional Fourier transform* ϕ and its inverse Fourier transform $\check{\phi}$ (if it exists) are given by

$$\begin{aligned}
\phi(\theta) &= \sum_{j\in\mathbb{Z}^d} f(j)e^{i\langle j,\theta\rangle}, \quad \theta\in[-\pi,\pi]^d \\
\check{\phi}(j) &= (2\pi)^{-d}\int_{[-\pi,\pi]^d}\phi(\theta)e^{-i\langle j,\theta\rangle}\,d\theta, \quad j\in\mathbb{Z}^d.
\end{aligned} \tag{1.3.7}$$

It is usually clear from the context which Fourier transform is appropriate, and we shall therefore often omit the word discrete.

For notational simplicity we shall describe the properties of Fourier transforms and the method of partial Fourier inversion in terms of functions F, $G\in L^1(\mathbf{R}^d)$. The analogous properties of the discrete case can easily be derived from this and are therefore not usually given. For details on d-dimensional transforms and the results quoted below, see Vo-Khac Khoan (1972).

A. The Inversion Theorem. If $F\in L^1(\mathbf{R}^d)$ and $\phi = FT(F)\in L^1(\mathbf{R}^d)$ and if

$$\check{\phi}(x) = (2\pi)^{-d}\int_{\mathbf{R}^d}\phi(\theta)e^{-i\langle x,\theta\rangle}\,d\theta \quad \text{for } x\in\mathbf{R}^d,$$

then $\check{\phi}$ is continuous and

$$F(x) = \check{\phi}(x) \quad a.e.$$

B. Conversion of convolutions to pointwise products. Let F, $G\in L^1(\mathbf{R}^d)$, and let ϕ and γ be their respective Fourier transforms. If H denotes the convolution of F and G and χ denotes its Fourier transform, then

$$\chi(\theta) = \phi(\theta)\gamma(\theta) \quad \text{for } \theta\in\mathbf{R}^d. \tag{1.3.8}$$

The next result constitutes one of the main theorems of Fourier transforms for L^2-functions. We shall make use of it repeatedly in Chapters 2 and 3 and therefore quote it for $L^2(\mathbf{R}^d)$ as well as for $\ell^2(\mathbb{Z}^d)$.

C. Parseval's identity. If F, $G\in L^2(\mathbf{R}^d)$, with Fourier transforms Φ and Γ, and f, $g\in\ell^2(\mathbb{Z}^d)$, with Fourier transforms ϕ and γ, then

$$\begin{aligned}
\int_{\mathbf{R}^d}F(x)\overline{G(x)}\,dx &= (2\pi)^{-d}\int_{\mathbf{R}^d}\Phi(\theta)\overline{\Gamma(\theta)}\,d\theta, \\
\sum_{j\in\mathbb{Z}^d} f(j)\overline{g(j)} &= (2\pi)^{-d}\int_{[-\pi,\pi]^d}\phi(\theta)\overline{\gamma(\theta)}\,d\theta.
\end{aligned} \tag{1.3.9}$$

1.3.2 Fourier Deconvolution

As we have seen in result **B** above, the Fourier transform interchanges convolution and pointwise product. This property of Fourier transforms is the key to Fourier filtering and Fourier deconvolution. We now turn to the blurred image B of (1.3.2).

Let τ, χ and β denote the Fourier transforms of T, H and B, respectively. Assume that $\chi(\theta) \neq 0$ for $\theta \in \mathbf{R}^d$. *Fourier deconvolution of B by χ* is the process which results in the function $\check{\tau}$ given by

$$\check{\tau}(x) = IFT(\beta\chi^{-1})(x) = (2\pi)^{-d} \int_{\mathbf{R}^d} \{\beta(\theta)/\chi(\theta)\} e^{-i\langle x,\theta\rangle} \, d\theta. \tag{1.3.10}$$

Because of its special rôle in Fourier deconvolution, the Fourier transform χ of the point spread function is called the *Fourier kernel.* In the engineering sciences, χ is often called a transfer function (see p212 of Rosenfeld and Kak (1976)); it may also be called a characteristic function if H is a probability density.

The inversion theorem **A** and (1.3.10) together show that T can be reconstructed a.e. from B, provided $\chi(\theta) \neq 0$. Problems arise however if $\chi(\theta) = 0$ for some $\theta \in \mathbf{R}^d$. Since $\beta = \chi\tau$, $\beta = 0$ whenever $\chi = 0$ and this leads to indeterminate ratios in the integrand of (1.3.10). Thus, even in the absence of noise, it is in general not possible to reconstruct T from B exactly. The situation is exacerbated for observations X which contain random noise ϵ. If ξ and ν denote the Fourier transforms of X and ϵ respectively, then the zeroes of ξ and ν will not usually coincide. As a consequence, ν/χ would be very much larger than τ in neighbourhoods of the zeroes of χ.

The problem of zeroes of χ in the deconvolution step can be overcome if one restricts the division by χ and the subsequent Fourier inversion to a subset of $\mathbf{R}^d$ on which χ is not too close to zero. This is the approach we adopt and now describe briefly.

Assume that $H \in L^1(\mathbf{R}^d)$. Let Θ be a subset of $\mathbf{R}^d$ on which χ^{-1} is defined and bounded. For observations $X = H \star T + \epsilon$, the *partial Fourier inversion estimator $\widehat{T}$ with smoothing set* Θ is defined by

$$\widehat{T}(x) = (2\pi)^{-d} \int_{\Theta} \{\xi(\theta)/\chi(\theta)\} e^{-i\langle x,\theta\rangle} d\theta. \tag{1.3.11}$$

The set Θ is also called the *inversion set.* Sometimes we want to emphasise which smoothing set is used in the definition of $\widehat{T}$. In this case we write $\widehat{T}_\Theta$ instead of $\widehat{T}$.

Often, the smoothing set will depend on a parameter, as in the following examples, which illustrate the main types of smoothing sets we shall employ.

For $\delta > 0$, put

$$\Theta_\delta = \{\theta \in \mathbf{R}^d : |\chi(\theta)| > \delta\}. \tag{1.3.12}$$

This choice of smoothing set is appropriate if $\chi(\theta) \to 0$ as $\theta \to \infty$. In this case, Θ_δ eliminates the high frequencies. These contribute very little in the case of smooth images T, and thus not much information is lost in employing Θ_δ instead of $\mathbf{R}^d$.

For functions χ with isolated zeroes at $\theta_i \in \mathbf{R}^d$, $i = 1, 2, \ldots$, a possible choice for Θ_δ is given by the following: Fix $\delta > 0$, and put

$$\begin{aligned} \Psi_\delta &= \{\psi \in \mathbf{R}^d : \|\psi - \theta_i\| < \delta \quad \text{for some } i\} \\ \Theta_\delta &= \mathbf{R}^d \setminus \Psi_\delta. \end{aligned} \tag{1.3.13}$$

Judicious choice of the inversion set Θ_δ is important as it provides a compromise between faithfulness of the estimator $\hat{T}$ to T and roughness of $\hat{T}$. As $\Theta_\delta \nearrow \mathbf{R}^d$ or as $\delta \to 0$ in (1.3.12)), $\hat{T} \to T + IFT(\nu\chi^{-1})$, and therefore $\hat{T}$ becomes less smooth as more frequency components are included in the inversion set. A smoother estimator $\hat{T}$ is obtained by selecting a smaller inversion set Θ_δ (e.g. by choosing a bigger value for the cut-off point δ); the noise component of $\hat{T}$ will be more smoothed out, while the image component of $\hat{T}$ becomes a worse approximation to T. Because of its rôle in the choice of the set Θ_δ, δ is often called the smoothing parameter, in analogy with smoothing parameters in statistical curve estimation.

The method of partial Fourier inversion has been used for a number of decades, especially in the engineering sciences where it is often referred to as Fourier filtering. In numerical analysis, on the other hand, the method is often described as numerical filtering (see for example Phillips (1962) and Twomey (1965)). Other authors (e.g. Hall (1987a), Titterington (1985)) call the estimator $\hat{T}$ of (1.3.11) a regularised estimator and associate zero weights with the set $\mathbf{R}^d \setminus \Theta_\delta$, and unit weights with Θ_δ. Practical implementations of Fourier deconvolution are described in Rosenfeld and Kak (1976) and in Koch and Tarlowski (1987).

1.4 Cross-Validation in Nonparametric Regression

1.4.1 Performance Measures and Rates of Convergence

Let Y_j denote data observed at points x_j, $j = 1, \ldots, n$ such that

$$Y_j = m(x_j) + \epsilon_j, \tag{1.4.1}$$

where m is an unknown function and ϵ denotes the observational error. Typically, in image analysis the unknown function m is the true image, and in regression m denotes the unknown regression function. In either case, one aims to find a function $\hat{m}$ which approximates m. To answer the question of what constitutes a reasonable approximation one usually considers pointwise or global distance measures. A common pointwise measure of accuracy is the mean square error, MSE, calculated at a fixed point x:

$$\mathrm{MSE}(x) = \mathbf{E}\{\hat{m}(x) - m(x)\}^2. \tag{1.4.2}$$

Sometimes global measures are preferable, since arguments that have been developed in the more general framework of global convergence also apply to problems of pointwise convergence. The following quadratic measures are often used:

$$\begin{aligned} \mathrm{SSE} &= \sum_{j=1}^{n} \{\hat{m}(x_j) - m(x_j)\}^2 \\ \mathrm{MSSE} &= \mathbf{E}(\mathrm{SSE}). \end{aligned} \tag{1.4.3}$$

In some cases, weights are part of the definition of SSE and MSSE. The measure SSE is called *sum of squared error* or *loss* and MSSE is called *mean sum of squared error* or *risk* (see p16 of Eubank (1988)). Many other choices of distance measure are available, such as the absolute and the maximum deviation of $\hat{m}$ and m, but we shall not be concerned with those. In Chapter 3 we shall use the global measures SSE and MSSE to assess the performance of cross-validation, while the simpler measure MSE of (1.4.2) will suffice when we consider the performance of partial Fourier inversion estimators in Chapter 2 and of median smoothers in Chapter 4.

In many situations one wants to assess the performance of an estimator as the sample size increases and to determine its rate of convergence to the true function m. Consider functions f, $g : \mathbb{N} \to \mathbb{R}_+$ which converge to zero as $n \to \infty$. We say f and g have *the same rate of convergence* if there exists a constant $c > 0$ such that $f \sim cg$.

Let d denote a distance measure and assume that d depends on the sample size n of the data (e.g. d =SSE or d =MSSE). Let $\hat{m} = \hat{m}(n)$ denote an estimator of m. We say that a function $r : \mathbb{R}_+ \to \mathbb{R}_+$ is an *optimal rate of convergence with respect to d* if there exist constants $c_1, c_2 > 0$ such that, as $n \to \infty$,

$$\begin{aligned} d\{\hat{m}(n), m\} &\leq c_1 r(n) \\ \inf_{\tilde{m}(n)} d\{\tilde{m}(n), m\} &\geq c_2 r(n), \end{aligned} \tag{1.4.4}$$

where the infimum is taken over all possible estimators for m. An estimator $\hat{m}$ of m is called *optimal with respect to d* if it satisfies the first of the two conditions (1.4.4).

In Chapter 2 we shall see how this definition of optimal rate of convergence can be adapted to a continuously defined parameter instead of the sample size n used above.

1.4.2 Choice of Smoothing Method

In practical applications of smoothing data one encounters two types of problems:

- the choice of the smoothing method;
- the choice of the smoothing parameter.

The selection of the smoothing method depends on a variety of factors, and different methods apply to regression data Y_j and indirect observations X_j of the form (1.1.2). For this reason, we shall only briefly review some of the methods used in nonparametric regression. As far as the second problem is concerned, the method of cross-validation for selecting the smoothing parameter will be described in Subsection 1.4.3, and it is this method which we shall generalise and adapt to the requirements of indirect observations X_j in Chapter 3.

In the context of nonparametric regression, the commonly used methods include local averaging, kernel smoothing and spline smoothing. In many of these methods, the estimator $\widehat{m}$ of m can be expressed as

$$\widehat{m}(x) = \sum \alpha_j(x) Y_j \tag{1.4.5}$$

for suitably chosen weights $\alpha_j(x)$. For kernel smoothing methods, Priestley and Chao (1972) were amongst the first to suggest the use of weights which are derived from densities and which can be parametrised by a scale parameter h. This parameter adjusts the size of the weights and is often called the *bandwidth* of the kernel weights. Typically one has weights $\alpha_j(x)$ given by

$$\alpha_j(x) = \kappa_h(x - x_j)/\hat{f}_h(x), \tag{1.4.6}$$

where

$$\begin{aligned}\hat{f}_h(x) &= n^{-1} \sum_{j=1}^{n} \kappa_h(x - x_j) \\ \kappa_h(u) &= h^{-1} \kappa(u/h).\end{aligned} \tag{1.4.7}$$

For details see Chapter 3 of Härdle (1989) and references therein.

In contrast to kernel smoothing, in the method of spline smoothing the estimator $\widehat{m}$ of m is taken to be the minimiser of S_λ over a class of admissible functions $\tilde{m}$, where

$$S_\lambda(\tilde{m}) = \sum_{j=1}^{n} \{Y_j - \tilde{m}(x_j)\}^2 + \lambda \int \{\tilde{m}^{(i)}(x)\}^2\, dx, \quad i = 0, 1, 2, \ldots. \tag{1.4.8}$$

$\tilde{m}^{(i)}$ denotes the ith derivative of $\tilde{m}$, and $\lambda > 0$. A comparison of (1.4.8) and (1.1.3) shows that spline smoothing is a special case of regularisation. The most common form of S_λ is that in which $i = 2$ in (1.4.8). For this case, Schoenberg (1964) and Reinsch (1967) have shown that if the class of admissible functions consists of all functions with square integrable second derivatives, then for each $\lambda > 0$, S_λ has a unique minimiser, which is a cubic spline. Furthermore, the minimiser of S_λ is of the form (1.4.5). A nice account of spline smoothing can be found in Wahba (1990).

The smoothing methods mentioned so far are distinguished by the existence of a well-developed theory and a wide range of applicability. Other smoothing techniques may be preferable under certain circumstances. For example, if one wanted an estimator which is resistant against outliers and is capable of modelling discontinuities, median smoothing would be more appropriate. In connection with blurred and noisy data $X_j = (k \star T)_j + \epsilon_j$ (see (1.1.2)), yet another smoothing method, namely partial Fourier inversion, is appropriate for finding an estimator $\widehat{T}$ of T. Common to all these methods, however, is the dependence of the estimator on a smoothing parameter.

1.4.3 Selection of the Smoothing Parameter by Cross-Validation

We assume that the distance measure SSE has been chosen to assess the performance of the estimator $\widehat{m}$. The estimator $\widehat{m}$ depends on a smoothing parameter h, which we indicate in this subsection by writing $\widehat{m}_h$. The aim now is to choose h in such a way that $\widehat{m}_h$ minimises the order of SSE. Such a choice of h will be called *optimal with respect to SSE.* Let $\hat{h}$ denote the optimal smoothing parameter. Then

$$\begin{aligned}\hat{h} &= \arg\min_h \mathrm{SSE}(h) \\ &= \arg\min_h \sum_{j=1}^{n} \{\widehat{m}_h(x_j) - m(x_j)\}^2 \\ &= \arg\min_h \left\{ \sum_{j=1}^{n} \widehat{m}_h(x_j)^2 - 2\sum_{j=1}^{n} \widehat{m}_h(x_j) m(x_j) \right\}, \qquad (1.4.9)\end{aligned}$$

since the term $\sum m(x_j)^2$ does not depend on h. This definition of the optimal h is useful for a theoretical analysis. However, in practice this approach cannot be used, since SSE cannot be computed without a knowledge of m itself. Closer inspection of (1.4.9) reveals that the term $\sum \widehat{m}_h(x_j)^2$ can be estimated from the data but the cross-term $\sum \widehat{m}_h(x_j)m(x_j)$ cannot. Since one cannot calculate this term from the data, the idea is to estimate it. A naive estimate would be $\sum \widehat{m}_h(x_j)Y_j$, where $m(x_j)$ has been replaced by Y_j. This procedure, however leads to a biased estimate of SSE (see Section 5.1 of Härdle (1989)), and is therefore not recommended.

Amongst the possible methods of finding an unbiased estimate of SSE, we concentrate on the leave-one-out or cross-validation technique (for some other methods see Section 5 of Härdle (1989)). The basic idea of cross-validation is the following:

- For each $j = 1, \dots, n$, divide the data into two uncorrelated parts such as Y_j and $\{Y_1, Y_2, \dots, Y_{j-1}, Y_{j+1}, \dots, Y_n\}$.
- Use one part to assess the performance of an estimator defined on the other.

For this leave-one-out method, one would estimate $m(x_j)$ by $\widetilde{m}(x_j) = Y_j$ and then compute an approximation $\widehat{m}_h^{\star}(x_j)$ of $\widehat{m}_h(x_j)$ from $\{Y_k : k \neq j\}$. Thus

$$\widetilde{R} \equiv \sum_{j=1}^{n} \widehat{m}_h^{\star}(x_j)^2 - 2\sum_{j=1}^{n} \widehat{m}_h^{\star}(x_j)\widetilde{m}(x_j) \tag{1.4.10}$$

would be an estimate for

$$R \equiv \sum_{j=1}^{n} \widehat{m}_h(x_j)^2 - 2\sum_{j=1}^{n} \widehat{m}_h(x_j)m(x_j). \tag{1.4.11}$$

In fact, cross-validation is a technique for selecting the smoothing parameter which is mathematically justified because one can show that under appropriate conditions the minimiser $\tilde{h}$ of $\widetilde{R}$ satisfies

$$\frac{\widetilde{R}(\tilde{h})}{\inf_h R(h)} \to 1 \quad a.s. \tag{1.4.12}$$

A proof of this for nonparametric regression can be found in Vieu (1991) and an earlier result, showing convergence in probability, is proved in Härdle and Marron (1985). For rates of convergence of $\tilde{h}$ to $\hat{h}$, see Härdle *et al.* (1988). For density estimation, an analogous result goes back to Stone (1984).

One of the features of cross-validation used in proving results of this kind is the fact that the cross-terms in $\widehat{m}_h^{\star}(x_j)\widetilde{m}(x_j)$ are calculated from uncorrelated parts of the data and the cross-term in $\widetilde{R}$ has therefore zero expectation. This observation is an important consideration in our development of cross-validation for blurred data in Chapter 3.

1.5 M-Estimation

For independent observations $Z_1, \dots, Z_n$ from a distribution with density f_θ, θ an unknown parameter, the maximum likelihood estimation strategy is to maximise the likelihood function, or equivalently, to minimise (over θ)

$$-\sum_{i=1}^{n} \log f_\theta(Z_i). \tag{1.5.1}$$

The minimiser $\hat{\theta}$ of (1.5.1) is called the maximum likelihood estimator of θ. For the estimation of the parameter in a location family (for which $f_\theta(z) = f(z-\theta)$), the idea of maximum likelihood estimation was generalised in Huber (1964) by allowing a larger class of functions in the definition (1.5.1). The resulting estimator is called an *M-estimator*, and is defined by

$$\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^{n} \rho(Z_i - \theta), \tag{1.5.2}$$

where ρ denotes a 'distance-like' function. The interest in a generalisation of maximum likelihood estimation was motivated by a desire to bound the influence of extreme or outlying observations, while at the same time maintaining acceptable efficiency when outliers are not present.

Often, the function ρ is differentiable and $\hat{\theta}$ is then a root of the equation

$$\sum_{i=1}^{n} \psi(Z_i - \theta) = 0, \tag{1.5.3}$$

where $\psi(z) = \rho'(z)$. If $\rho(z) = z^2$, for example, solution of (1.5.3) leads to $\hat{\theta} = \bar{Z}$, the sample mean. As well as the mean, there are many other useful M-estimators which can be determined by choosing ψ in (1.5.3) (see p103ff of Hampel (1986)). In cases more complex than the mean, (1.5.3) must be solved numerically. Numerical methods for solving equations often rely on the existence of the derivative, and, similarly, theoretical properties of the estimator are also often derived under the assumption that ψ has a derivative.

The median is an M-estimator with $\rho(z) = |z|$. The function ψ as in (1.5.3) exists for the median, namely, $\psi(z) = \mathrm{sign}(z)$, and the median is a root of (1.5.3), which can be found explicitly. However, because of the non-differentiability of this ψ, theoretical properties of the median are more difficult to obtain.

There are many natural generalisations of these ideas to the estimation of parameters in more general models. We shall be concerned in particular with nonparametric regression problems in which one wants to estimate locally the location of the dependent variable. To estimate the regression function at the point x by local averaging, one averages the values of the dependent variable corresponding to design points in a suitably chosen neighbourhood of x. The influence of outliers on the mean is not bounded, so this procedure can lead to poor estimates if there is outlier contamination. It is this observation that leads to the idea of replacing the local average by a local location M-estimator which is more resistant to outliers. This approach was analysed by Härdle and Gasser (1984). In Chapter 4, we extend their results, which were derived under the assumption that the function corresponding to ψ in (1.5.3) was differentiable, to include the case of the local median, for which the corresponding function ψ is not differentiable.

Chapter 2

Partial Fourier Inversion

2.1 Introduction

Most theoretical analyses of indirect images are confined to discrete observation points. This clearly reflects the practical set-up: data are recorded at discrete points in time or space. For the latter, lattices based on squares or rectangles are mostly employed, but some imaging devices such as texture analysers (see Serra (1982)) also use triangular and hexagonal lattices.

For a theoretical study of images, however, the discrete nature of the recorded data corresponds to modelling the sampling distribution, i.e. the point spread function, by a discrete distribution. This imposes severe limitations on the class of distributions that can be treated (see Hall (1990)), and excludes interesting and realistic distributions,such as those treated below.

For this reason, the analysis presented here is based on continuously defined observations. This situation may be regarded as the limiting case of discretely observed data when the pixel size or the separation between observation points tends to zero. Typically we shall assume that X is observed at points in a convex subset $\mathcal{R}$ of $\mathbf{R}^d$. This will allow us to treat a large class of problems simultaneously, including the out-of-focus blur and the motion blur of Hall (1990).

In the models discussed in this chapter, we shall always assume that the point spread function is known. This corresponds to knowing (or measuring) the effect of the imaging apparatus, and is common practice in many applications in engineering.

All our observations are of the form

$$X = H \star T + N, \tag{2.1.1}$$

where $\star$ denotes convolution, T denotes the true image, H the known point spread function and N denotes random noise. We examine the method of partial Fourier inversion and show that it provides an optimal method for image enhancement for a wide range of images and point spread functions. Fourier inversion techniques and, in particular, the Wiener filter are regarded as standard tools in the engineering sciences. We concentrate on deterministic images, while the Wiener filter approach assumes the images to be weakly stationary random fields. For this reason the Wiener filter does not apply to the situation considered here.

In Hall (1987a, 1987b) motion blur and out-of-focus blur were applied to piecewise constant images or test-patterns and the performance of restoration methods was considered from the viewpoint of consistency. Somewhat later, in Hall (1990), the same point spread functions were used to degrade images which may be regarded as discretisations of continuously differentiable functions, and precise upper and lower bounds for the mean square error were calculated. The results presented here extend those of Hall (1990) in several important aspects:

- By considering continuously defined image models a wide range of realistic point spread functions can be treated in a systematic way.
- Dependent noise models are developed which replace the assumption of white noise treated in Hall (1990). Although much simpler to deal with, the white noise model is less general.

An abridged version of the results described here is given in Hall and Koch (1990).

In the next section we describe the class of images and point spread functions that will be considered. Section 2.3 is concerned with models for correlated noise, and considers moving average noise, which may be taken as a concrete example of the kind of noise we are dealing with. In Section 2.4 we describe our results, in particular our main result, Theorem 2.6, which shows that under suitable regularity conditions, partial Fourier inversion is an optimal reconstruction method in a mean square sense. Section 2.5 contains some simple examples which demonstrate the scope of our theorem. Furthermore, we look at some applications which give specific rates of convergence for the partial Fourier inversion estimator. Proofs of the results of Sections 2.4 and 2.5 appear in Section 2.6.

2.2 Models for Images and Image Degradation

In this section we give a description of classes of images and blur which will form the framework for Theorem 2.6. We shall also consider specific examples of images and blur functions. These will lead to specific rates of convergence as well as demonstrate the necessity of the assumptions in Theorem 2.6.

2.2.1 Image Models

We assume that the true images T are real-valued functions, defined on $\mathbf{R}^d$ ($d \in \mathbf{N}$, $d \geq 1$), which are absolutely integrable. These properties of T we summarise by writing $T \in L^1(\mathbf{R}^d, \mathbf{R})$. The integrability of T guarantees the existence of a continuous and bounded Fourier transform τ of T. In fact, it is convenient to describe classes of images in terms of their Fourier transforms in the following way.

For $\tau_0 \in L^1(\mathbf{R}^d, \mathbf{R}_+)$, τ_0 symmetric (i.e. $\tau_0(\theta) = \tau_0(-\theta)$), define the image class $\mathcal{C}(\tau_0)$ by

$$\mathcal{C}(\tau_0) \equiv \{T \in L^1(\mathbf{R}^d, \mathbf{R}) \ : \ |\tau(\theta)| \leq \tau_0 \ \forall \theta \in \mathbf{R}^d\} \tag{2.2.1}$$

and call τ_0 the *envelope* of the class of images $\mathcal{C}(\tau_0)$. Thus τ_0 is an envelope if $\tau_0 \in L^1(\mathbf{R}^d, \mathbf{R}_+)$, τ_0 is symmetric and τ_0 defines a class of images in $L^1(\mathbf{R}^d, \mathbf{R})$. Let T_0 denote the inverse Fourier transform of τ_0. If $T_0 \in L^1(\mathbf{R}^d, \mathbf{R})$, then $T_0 \in \mathcal{C}(\tau_0)$.

In this chapter, τ_0 will always denote an envelope of a class of images. From (2.2.1) it follows that the class of images $\mathcal{C}(\tau_0)$ is completely characterised by properties of τ_0. Of particular interest here will be those envelopes τ_0 of polynomial decay and of exponential decay, described next. Locally, images of this type may be regarded as approximations to realistic images such as extended brightness sources in astronomical imaging.

A. Images with polynomially decreasing Fourier transforms. Assume that there exist $A > 0$, $a > 0$ such that

$$\tau_0(\theta) = A(1 + \|\theta\|)^{-a} \quad \text{for } \theta \in \mathbf{R}^d. \tag{2.2.2}$$

Equation (2.2.2) implies that $\sup_\theta\{\tau_0(\theta)\|\theta\|^a\} < \infty$. If furthermore $M^k\tau_0 \in L^1(\mathbf{R}^d)$, for $k \leq a$, where $(M^k\tau_0)(\theta) = \theta^k\tau_0(\theta)$, then

$$\mathcal{C}(\tau_0) = \{T \in L^1(\mathbf{R}^d, \mathbf{R}) \ : \ |\tau(\theta)| \leq A(1 + \|\theta\|)^{-a} \ \forall \theta \in \mathbf{R}^d\}$$

corresponds to the set of true images which are a times continuously differentiable, that is, $\mathcal{C}(\tau_0) \subseteq C^a(\mathbf{R}^d)$ (see p.26 théorème of Vo-Khac Khoan (1972)). Here the parameter a indicates the smoothness class of the images. The higher the value of a, the smoother the images are. Images of this kind are used in Examples 1 and 3 of Subsection 2.5.1.

B. Images with exponentially decreasing Fourier transforms. For constants $A > 0$, $a > 0$, define τ_0 by

$$\tau_0(\theta) = A\exp(-a\|\theta\|_p^p) \quad \text{for } \theta \in \mathbf{R}^d, \quad p \geq 1. \tag{2.2.3}$$

This type of image is used in Examples 2 and 4 of Subsection 2.5.1. Of special interest are the cases $p = 1, 2$, which we now consider.

If $p = 2$, then $\tau_0(\theta) = A \prod_{i=1}^{d} \exp(-a|\theta_i|^2)$, and $T_0 = IFT(\tau_0)$ is given by

$$T_0(x) = A \exp(-\|x\|^2/4a) \quad \text{for } x \in \mathbf{R}^d. \tag{2.2.4}$$

The class $\mathcal{C}(\tau_0)$ then is a subset of $C^\infty(\mathbf{R}^d, \mathbf{R})$, the class of infinitely often differentiable functions from $\mathbf{R}^d$ to $\mathbf{R}$:

$$\mathcal{C}(\tau_0) = \{T \in C^\infty(\mathbf{R}^d, \mathbf{R}) \ : \ |T(x)| \leq A \exp(-\|x\|^2/4a) \ \forall x \in \mathbf{R}^d\}.$$

If $p = 1$, then $T_0 = IFT(\tau_0)$ is given by

$$T_0(x) = A \prod_{i=1}^{d} \{2a/(x_i^2 + a^2)\} \quad \text{for } x \in \mathbf{R}^d, \tag{2.2.5}$$

and $\mathcal{C}(\tau_0)$ corresponds to the set of images which are proportional to products of Cauchy distributions (see also paragraph **A** of Subsection 2.2.2 below).

We could easily treat more general classes of images than those given in paragraphs **A** and **B** above, such as combinations of (2.2.2) and (2.2.3) or images with envelopes τ_0 of the form $\tau_0(\theta) = A(1 + \|\theta\|)^{-a}\{1 + \max(0, \log\|\theta\|)\}^{-b}$, $A, a, b > 0$. For reasons of clarity and simplicity, the examples given in later sections will concentrate on images of the classes described above .

2.2.2 Models for the Blur

In our model, the true image cannot be observed directly, but is degraded in a linear and systematic way by a deterministic function. This function is assumed to be known and will be denoted by H throughout this chapter. It is called the blur or the point spread function; its Fourier transform χ is called the *Fourier kernel.*

We assume that

$$H : \mathbf{R}^d \to \mathbf{R} \text{ is symmetric, and } \int H(x)\, dx = 1.$$

The assumption $\int H = 1$ allows us to regard H as a density, and its Fourier transform χ as a characteristic function, as will be convenient in paragraph **A** below.

The *blurred image*, denoted by B, is of the form

$$B = H \star T.$$

This image would be observed in the absence of random noise.

A. Point spread functions with thin tails. We are interested in point spread functions H which are concentrated near the origin, that is, point spread functions with compact support or with "thin" tails .

Polynomially decreasing kernels: For $a > 0$, define χ by

$$\chi(\theta) = c(1 + \|\theta\|)^{-a} \text{ for } \theta \in \mathbf{R}^d, \tag{2.2.6}$$

where $c > 0$ is chosen such that $\chi \in L^1(\mathbf{R}^d)$ and $\int H(x)dx = 1$. If $M^a\chi \in L^1(\mathbf{R}^d)$, where $(M^a\chi)(\theta) = \theta^a\chi(\theta)$ for $\theta \in \mathbf{R}^d$, then H is an a times continuously differentiable function. (See also paragraph A of Subsection 2.2.1 and Examples 1 and 4 in Subsection 2.5.1.)

Exponentially decreasing kernels: For $c > 0, \alpha > 0$ define χ by

$$\chi(\theta) = \exp(-c\|\theta\|^\alpha) \text{ for } \theta \in \mathbf{R}^d. \tag{2.2.7}$$

Fourier kernels of this kind are also used in Examples 2 and 3 in Subsection 2.5.1.

Of particular practical interest is the case $\alpha = 1$ and $d=2$. Fourier kernels of this kind appear, for example, in remote sensing problems. For $\alpha = 1$ the blur function H becomes

$$H(x) \propto c\{\|x\|^2 + c^2\}^{-3/2} \quad \text{for } x \in \mathbb{R}^2; \tag{2.2.8}$$

which may be interpreted as the filter corresponding to downward continuation by distance c in the magnetic field equation (see Koch and Tarlowski (1987)).

Characteristic functions of stable distributions: For $0 < \alpha \leq 2$, $c > 0$, define χ by:

$$\chi(\theta) = \exp(-c\sum_{i=1}^{d} |\theta_i|^\alpha) \text{ for } \theta \in \mathbf{R}^d. \tag{2.2.9}$$

The corresponding point spread function H is given by

$$H(x) = (2\pi)^{-d} \prod_{j=1}^{d} \int_{\mathbf{R}} \exp(-c|\theta_j|^\alpha) \exp(-i\theta_j x_j)\, d\theta_j \text{ for } x \in \mathbf{R}^d. \tag{2.2.10}$$

By Theorem 2.2.2, p43, of Ibragimov and Linnik (1971), it follows that χ is the d-dimensional product of characteristic functions of a stable distribution. Such distributions are of interest in their own right, as they are, under weak assumptions, the only possible limiting distributions of normal sums of stationary dependent variables.

The point spread function H given in (2.2.10) is a symmetric function which has derivatives of all orders for $x \in \mathbf{R}^d$. Expressions for H for $\alpha = 1$ and $\alpha = 2$ were given in (2.2.5) and (2.2.4) respectively. For $\alpha = 2$, the Fourier kernel $\chi(\theta) = \exp(-c||\theta||^2)$ is also a special case of (2.2.7). A further description of the densities of stable distributions for $\alpha \neq 1$, $\alpha \neq 2$ can be found in Ibragimov and Linnik (1971), Theorem 2.3.1, p48.

B. Point spread functions with compact support. We describe two kinds of point spread functions here: products and convolution products of functions with compact support.

Products: For $\nu \in \mathbf{N}$, $\nu \geq 1$, define H_ν on $\mathbf{R}^d$ by:

$$H_\nu(x) = \begin{cases} C_1(\nu)\prod_{i=1}^d \{\cos(\pi x_i/2)\}^{\nu-1} & \text{if} \quad x_i \in [-1,+1] \\ 0 & \text{if} \quad x_i \in \mathbf{R}\setminus[-1,+1], \end{cases} \tag{2.2.11}$$

where $C_1(\nu) = \left[\int_{-1}^{+1}\{\cos(\pi x/2)\}^{\nu-1}\,dx\right]^{-d}$, so $\int H = 1$.

For $\nu = d = 1$, H_1 is just motion blur. As ν increases, H_ν becomes smoother near 1, and more concentrated near 0, becoming the d-dimensional Dirac delta-function in the limit.

The Fourier kernel χ_ν on $\mathbf{R}^d$ is given by

$$\begin{aligned} \chi_\nu(\theta) &= \{2^{1-\nu}\pi^{-1}\Gamma(\nu)\}^d C_1(\nu) \\ &\quad \times \prod_{i=1}^d \sin(\theta_i - \frac{\nu-1}{2}\pi)\Gamma(\frac{\theta_i}{\pi} - \frac{\nu-1}{2})\left\{\Gamma(\frac{\theta_i}{\pi} + \frac{\nu+1}{2})\right\}^{-1}. \end{aligned} \tag{2.2.12}$$

This follows from Lemma 2.6.2, which is proved in Subsection 2.6.4. This lemma also gives the zeroes of χ_ν for the case $d = 1$, which occur at $\theta = \pm(\nu-1)\pi/2 + n\pi$, $n \geq 1$. Since χ_ν is the d-dimensional product of the one-dimensional version, $\chi_\nu = 0$ on hyperplanes $\mathcal{N}_\nu$ given by

$$\mathcal{N}_\nu = \{\theta = (\theta_i) : \theta_i = \pm(\nu-1)\pi/2 + n\pi, n \geq 1 \text{ for some } i = 1,\ldots,d\}$$

Furthermore, as $||\theta|| \to \infty$,

$$|\chi_\nu(\theta)| \asymp \prod_{i=1}^d \left|\sin(\theta_i - \frac{\nu-1}{2}\pi)\right|(1+|\theta_i|)^{-\nu},$$

thus χ_ν decreases as $||\theta||$ increases, as well as containing zeroes.

Convolution products: Let Π denote the rectangular function in $\mathbf{R}^d$ defined by

$$\Pi(x) = \begin{cases} 2^{-d} & x \in [-1,+1]^d \\ 0 & x \in \mathbf{R}^d \setminus [-1,+1]^d. \end{cases}$$

For $\nu \in \mathbf{N}$, $\nu \geq 1$, let H_ν denote the ν-fold convolution product of Π:

$$H_\nu(x) = c_1(\nu) \star_\nu \Pi(x), \tag{2.2.13}$$

where $c_1(\nu)$ is chosen so that $\int H_\nu = 1$.

As in the case of (2.2.8), for $\nu = d = 1$, H_1 is motion blur, and for $\nu = 2$, H_ν is the product of triangles with support $[-2,2]$. As ν increases, H_ν becomes smoother, and its support increases.

If Π_1 denotes the rectangular function on $[-1,1]$ of height $1/2$, whose Fourier transform is $\theta^{-1}\sin\theta$, then $\Pi(x) = \prod_{i=1}^d \Pi_1(x_i)$, and thus the Fourier kernel χ_ν of H_ν becomes

$$\chi_\nu(\theta) = c_1(\nu) \prod_{l=1}^{\nu} \left(\prod_{i=1}^{d} \frac{\sin\theta_i}{\theta_i} \right) = c_1(\nu) \prod_{i=1}^{d} \left(\frac{\sin\theta_i}{\theta_i} \right)^\nu, \quad \theta \in \mathbf{R}^d. \tag{2.2.14}$$

As in the previous example, $\chi_\nu = 0$ on hyperplanes $\mathcal{N}$ where

$$\mathcal{N} = \{\theta = (\theta_i) : \theta_i = n\pi, n \in \mathbb{Z} \text{ for some } i = 1, \ldots, d\},$$

and, asymptotically, χ_ν decreases polynomially:

$$\chi_\nu(\theta) \sim \prod_{i=1}^{d} (\sin\theta_i)^\nu (1 + |\theta_i|)^{-\nu} \quad \text{as } \|\theta\| \to \infty.$$

2.3 Models for Correlated Noise

The random degradation in the observations is modelled by an additive random field $N : \mathbf{R}^d \to \mathbf{R}$. As we are deviating from the white noise model, it will be convenient to characterise properties of the noise process N in terms of the *autocovariance function* γ_N of N defined by

$$\gamma_N(x,y) = \mathbf{E}\left[\{N(x) - \mathbf{E}N(x)\}\{N(y) - \mathbf{E}N(y)\}\right] \quad \text{for } x, y \in \mathbf{R}^d. \tag{2.3.1}$$

The process N is *(second order) stationary* if its autocorrelation function γ_N satisfies

$$\gamma_N(x,y) = \gamma_N(x+z, y+z) \quad \text{for } x, y, z \in \mathbf{R}^d. \tag{2.3.2}$$

If γ_N satisfies (2.3.2), it can easily be seen that γ_N depends only on the difference $x - y$, and may therefore be regarded as a function $\gamma_N : \mathbf{R}^d \to \mathbf{R}$, as is done from now on.

Condition (2.3.2) is sometimes called *weak stationarity* or *covariance stationarity.* As a consequence of (2.3.2), γ_N is symmetric, $\gamma_N(0) \geq 0$, and γ_N attains its maximum at $x = 0$. (For further properties of γ_N see p11ff of Brockwell and Davis (1987).)

Assume that the random field $N : \mathbf{R}^d \to \mathbf{R}$ satisfies

N1 $\mathbf{E}N = 0, \quad \mathbf{E}N^2 = 1;$

N2 N is (second order) stationary;

N3 $\gamma_N \in L^1(\mathbf{R}^d)$.

The last property of N guarantees the existence of the Fourier transform f_N of γ_N, called the *spectral density,* which is defined by

$$f_N(\theta) = \int_{\mathbf{R}^d} \gamma_N(x) e^{i\langle x,\theta\rangle}\, dx \quad \text{for } \theta \in \mathbf{R}^d. \tag{2.3.3}$$

The bounds on the mean square error discussed in the next section depend on the noise solely via its spectral density.

For the remainder of this chapter, N will always denote a second order stationary random field, satisfying **N1–N3**.

Let $\gamma_N = \{\gamma_{N,t} : t > 0\}$ denote a family of autocovariance functions for N. The value of t governs the amount of dependence or the range of dependence of the noise process N: we assume that the dependence between $N(x)$ and $N(y)$ decreases as t decreases.

As an illustration, let γ_N denote the autocovariance function of N and define the family $\{\gamma_{N,t}\}$ by:

$$\left.\begin{array}{rcl} \gamma_{N,1}(x) & = & \gamma_N(x) \\ \gamma_{N,t}(x) & = & \gamma_N(t^{-1}x) \end{array}\right\} \quad \text{for } x \in \mathbf{R}^d, \quad t > 0. \tag{2.3.4}$$

As $t \to 0$, $\gamma_{N,t}$ tends to zero more quickly. It follows that the correlation between $N(x)$ and $N(y)$ diminishes more rapidly too.

A family of autocovariance functions $\{\gamma_{N,t}\}$ gives rise to a corresponding family of spectral densities $\{f_{N,t}\}$ via the Fourier transform relationship (2.3.3). If $\gamma_{N,t}$ is defined by (2.3.4), then $f_{N,t}$ satisfies:

$$\left.\begin{array}{rcl} f_{N,1}(\theta) & = & f_N(\theta) \\ f_{N,t}(\theta) & = & t^d f_N(t\theta) \end{array}\right\} \quad \text{for } \theta \in \mathbf{R}^d, t > 0. \tag{2.3.5}$$

We are now in a position to state further properties of the noise process that will be required in Theorem 2.6. Let $f_N = \{f_{N,t}\}$ denote a family of spectral densities of N such that

N4 $|f_{N,t}(\theta)| \leq c_1 t^d$ for $\theta \in \mathbf{R}^d$;

N5 $\inf_{||\theta|| \leq c_2 t^{-1}} f_{N,t}(\theta) \geq c_3 t^d$ for $\theta \in \mathbf{R}^d$;

for constants c_1, c_2, $c_3 > 0$, where c_2 may depend on c_3.

It is easy to check that the family $\{f_{N,t}\}$ given by (2.3.5) satisfies **N5**, if f_N is bounded away from zero in a neighbourhood of $\theta = 0$, since

$$\inf_{||\theta|| \leq c_2 t^{-1}} f_{N,t}(\theta) = \inf_{||t\theta|| \leq c_2} t^d f_N(t\theta).$$

Note that f_N is bounded, being the Fourier transform of γ_N, and thus **N4** holds.

For the remainder of this section we look at a specific example of a correlated noise process which satisfies **N1–N5**.

Let W denote a d-dimensional Wiener process or Brownian sheet, that is, a Gaussian process defined on $\mathbf{R}^d_+$ with mean zero and covariance structure

$$\mathbf{E}\{W(x)W(y)\} = \prod_{i=1}^{d}(x_i \wedge y_i), \tag{2.3.6}$$

where $u \wedge v = \min(u, v)$. For more details on Wiener processes see Adler (1981).

For $l > 0$, put $A_l = [0, l]^d$, and define $N : \mathbf{R}^d \to \mathbf{R}$ by

$$N(x) = \int_{\mathbf{R}^d} b(u)\, dW(x - u), \tag{2.3.7}$$

where $b(x) = l^{-d/2}$ if $x \in A_l$ and $b(x) = 0$ otherwise, and W is a Wiener process as in (2.3.6).

Proposition 2.1 *Assume that N satisfies (2.3.7). Then N is an MA(l)-process with spectral density f_N given by*

$$f_N(\theta) = l^d \prod_{i=1}^{d} \sin^2(l\theta_i/2)/(l\theta_i/2)^2 \quad \textit{for } \theta \in \mathbf{R}^d \tag{2.3.8}$$

which satisfies **N1–N5**.

Proof of Proposition 2.1

Since N satisfies (2.3.7), N is given by a generalised Stieltjes integral in the sense of Doob (1953). It follows from Theorem 2.1 and Sections 2 to 4 of Chapter

IX of Doob that N is a random process such that $\mathbf{E}N = 0$. By Section 8 Chapter XI of Doob, N may be regarded as a moving average process, and is therefore stationary. A change of variable now allows us to rewrite N in the following form:

$$N(x) = \int_{x-l\mathbf{j}}^{x} b(x-u)\, dW(u), \tag{2.3.9}$$

where $\mathbf{j} = (1, 1, \ldots, 1)^T \in \mathbf{R}^d$. Now, for x, $\tau \in \mathbf{R}^d$, $\gamma_N(\tau) = \mathbf{E}\{N(x+\tau)N(x)\}$ is given by

$$\begin{aligned} \mathbf{E}\{N(x+\tau)N(x)\} &= \int_{x+\tau-l\mathbf{j}}^{x+\tau}\int_{x-l\mathbf{j}}^{x} b(x+\tau-u)b(x-v)\mathbf{E}\{dW(u)dW(v)\} \\ &= \int_{x+\tau-l\mathbf{j}}^{x+\tau}\int_{x-l\mathbf{j}}^{x} b(x+\tau-u)b(x-v)\delta(u-v)\, dudv. \end{aligned}$$

Here δ denotes the Dirac delta-function in d dimensions, and

$$\delta(u-v)dudv = \mathbf{E}\{dW(u)dW(v)\},$$

which follows from (2.3.6). Next observe that $\int\int \delta(u-v)\, dudv = \int dv$, and thus

$$\begin{aligned} \mathbf{E}\{N(x+\tau)N(x)\} &= \int_{x-l\mathbf{j}}^{x} b(x+\tau-v)b(x-v)\, dv \\ &= \int_{A_l} b(v-\tau)b(v)\, dv \\ &= \begin{cases} l^{-d}\prod_{i=1}^{d}(l-|\tau_i|) & \text{for } \tau \in A_l \\ 0 & \text{for } \tau \notin A_l. \end{cases} \end{aligned}$$

~~This shows that N is second order stationary.~~ Thus we may write

$$\gamma_N(\tau) = \mathbf{E}\{N(x+\tau)N(x)\} = l^{-d}\prod_{i=1}^{d}(l-|\tau_i|)\mathcal{I}_{A_l} \tag{2.3.10}$$

where $\mathcal{I}_{A_l}$ denotes the indicator function of A_l. If $\tau = 0$, $\gamma_N(0) = 1$ by (2.3.10), showing that N has unit variance for each $x \in \mathbf{R}^d$. Clearly, $\gamma_N \in L^1(\mathbf{R}^d)$.

Since N is stationary and W is a Gaussian process, (2.3.7) defines a d-dimensional MA(l)-process.

It remains to show (2.3.8), N4 and N5. Here we use the fact that f_N, the spectral density, can be written as the d-fold product of one-dimensional Fourier transforms. For $\theta \in \mathbf{R}^d$,

$$\begin{aligned} f_N(\theta) &= \int_{\mathbf{R}^d} \gamma_N(x)e^{i\langle x,\theta\rangle}\, dx \quad \text{(by (2.3.3))} \\ &= l^{-d}\int_{A_l}\prod_{i=1}^{d}(l-|x_i|)\exp(i\Sigma x_i\theta_i)\, dx \\ &= l^{-d}\prod_{i=1}^{d}\{\sin(l\theta_i/2)/(\theta_i/2)\}^2 \\ &= l^{d}\prod_{i=1}^{d}\{\sin(l\theta_i/2)/(l\theta_i/2)\}^2. \end{aligned} \tag{2.3.11}$$

Insert l-10 on p30

$$\begin{aligned} T_\Theta(x) &= (2\pi)^{-d} \int_\Theta \tau(\theta) e^{-i\langle x,\theta\rangle}\, d\theta \\ k_\Theta(x) &= (2\pi)^{-d} \int_\Theta \chi(\theta)^{-1} e^{-i\langle x,\theta\rangle}\, d\theta. \end{aligned} \tag{2.4.3}$$

The limit (2.4.2) exists under the assumptions that N has continuous sample paths and that $\chi^{-1}\mathcal{I}_\Theta \in L^1(\mathbb{R}^d)$. To see this, note that

$$\begin{aligned} \widehat{T}_\mathcal{R}(x) &= (2\pi)^{-d} \int_\Theta \int_{\mathbb{R}^d} (H \star T)(y)\mathcal{I}_\mathcal{R}(y)\chi(\theta)^{-1}\mathcal{I}_\Theta(\theta) e^{-i\langle y-x,\theta\rangle}\, dy d\theta \\ &\quad +(2\pi)^{-d} \int_\Theta \int_\mathcal{R} N(y)\chi(\theta)^{-1} e^{-i\langle y-x,\theta\rangle}\, dy d\theta \end{aligned}$$

For the deterministic part of $\widehat{T}_\mathcal{R}$, the integrand is dominated by

$$g_1(y,\theta) \equiv |(H\star T)(y)\chi(\theta)^{-1}\mathcal{I}_\Theta(\theta)|,$$

and $g_1 \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ by Problem 1, p106 of Dudley (1989). The Dominated Convergence Theorem therefore applies to yield

$$\begin{aligned} &\lim_{\mathcal{R}\uparrow\mathbb{R}^d} \int_\Theta \int_{\mathbb{R}^d} (H \star T)(y)\mathcal{I}_\mathcal{R}(y)\chi(\theta)^{-1}\mathcal{I}_\Theta(\theta) e^{-i\langle y-x,\theta\rangle}\, dy d\theta \\ &= (2\pi)^{-d} \int_\Theta \int_{\mathbb{R}^d} \lim_{\mathcal{R}\uparrow\mathbb{R}^d} (H \star T)(y)\mathcal{I}_\mathcal{R}(y)\chi(\theta)^{-1}\mathcal{I}_\Theta(\theta) e^{-i\langle y-x,\theta\rangle}\, dy d\theta \\ &= (2\pi)^{-d} \int_\Theta \tau(\theta) e^{i\langle x,\theta\rangle}\, d\theta = T_\Theta(x). \end{aligned} \tag{2.4.3a}$$

Let $\mathrm{ran}_\mathcal{R}$ denote the random part of $\widehat{T}_\mathcal{R}$. It follows that

$$\mathrm{ran}_\mathcal{R}(x) = \int_\mathcal{R} N(y) k_\Theta(x-y)\, dy \tag{2.4.4}$$

by (2.4.3) and the fact that $\mathcal{R}$ is bounded and N has continuous sample paths. Equation (2.4.4) defines a mean zero, finite variance random varible which has continuous and bounded sample paths. The variance of $\mathrm{ran}_\mathcal{R}$ is given by

$$\begin{aligned} \mathrm{var}\ \mathrm{ran}_\mathcal{R}(x) &= \int_{\mathcal{R}\times\mathcal{R}} \mathbb{E}\{N(x-u)N(x-v)\} k_\Theta(u)k_\Theta(v)\, dudv \\ &= \int_{\mathcal{R}\times\mathcal{R}} \gamma_N(u-v) k_\Theta(u)k_\Theta(v)\, dudv. \end{aligned}$$

As $\mathcal{R} \uparrow \mathbb{R}^d$,

$$\int_\mathcal{R} \gamma_N(u-v)k_\Theta(u)\, du \to \int_{\mathbb{R}^d} \gamma_N(u-v)k_\Theta(u)\, du = (\gamma_N \star k_\Theta)(v)$$

by Theorem 7.14 of Rudin (1970). Since $\gamma_N \star k_\Theta \in L^1(\mathbb{R}^d)$, Proposition 5.1.13 of Butzer and Nessel (1971) yields that

$$\mathrm{var}\ \mathrm{ran}_\mathcal{R}(x) \to \int_{\mathbb{R}^d\times\mathbb{R}^d} \gamma_N(u-v)k_\Theta(u)k_\Theta(v)\, dudv < \infty.$$

This shows that $\lim_{\mathcal{R}\uparrow\mathbb{R}^d} \mathrm{var}\ \mathrm{ran}_\mathcal{R}(x)$ exists.

Next let $\mathcal{R}_n$ denote a nested sequence of bounded increasing subsets of $\mathbb{R}^d$. The corresponding sequence of random variables $\mathrm{ran}_{\mathcal{R}_n}(x)$ forms a Cauchy sequence with respect to mean square, so $\mathrm{ran}_{\mathcal{R}_n}(x) \in L^2(d\mathbb{P})$, where $\mathbb{P}$ denotes the probability measure corresponding to the expectation $\mathbb{E}$. But $L^2(d\mathbb{P})$ is complete, and therefore the mean square limit $\mathrm{ran}(x)$ belongs to $L^2(d\mathbb{P})$. Together with (2.4.3a) this shows that $\widehat{T}(x)$ exists, and we may therefore write

$$\widehat{T}(x) = T_\Theta(x) + \lim_{\mathcal{R}\uparrow\mathbb{R}^d} \int_\mathcal{R} N(y)k_\Theta(x-y)\, dy, \tag{2.4.5}$$

Putting $f_N(\theta) = \{f_{N,l}(\theta)\}$ it follows from (2.3.11) that N4 and N5 are satisfied for $t = l$, $c_1 = 1$, $c_2 = \pi$ and $c_3 = (2/\pi)^{2d}$. □

2.4 Main Results

2.4.1 Bounds for the Mean Square Error

For observations $X = H \star T + N$, we now look at the performance of estimators $\widetilde{T}$ of T, measured by the mean square error of $\widetilde{T}$. Of special interest is the partial Fourier inversion estimator $\widehat{T}$ of T (see (1.3.11)).

We assume that the data X are recorded in a bounded region $\mathcal{R} \subseteq \mathbf{R}^d$. Let $\xi_{\mathcal{R}}$ denote the Fourier transform of the recorded data, that is,

$$\xi_{\mathcal{R}}(\theta) = \int_{\mathcal{R}} X(x) e^{i\langle x,\theta\rangle}\, dx \quad \text{for } \theta \in \mathbf{R}^d, \tag{2.4.1}$$

and put

$$\widehat{T}_{\mathcal{R}}(x) = (2\pi)^{-d} \int_{\Theta} \{\xi_{\mathcal{R}}(\theta)/\chi(\theta)\} e^{-i\langle x,\theta\rangle}\, d\theta \quad \text{for } x \in \mathbf{R}^d,$$

where Θ denotes a symmetric smoothing set for χ, such as the set Θ_δ defined in (1.3.12). The partial Fourier inversion estimator $\widehat{T}$ of T is then given by

$$\widehat{T}(x) = \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \widehat{T}_{\mathcal{R}}(x) = (2\pi)^{-d} \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \int_{\Theta} \{\xi_{\mathcal{R}}(\theta)/\chi(\theta)\} e^{-i\langle x,\theta\rangle}\, d\theta, \tag{2.4.2}$$

where the limit is taken in the mean square sense. Put Insert p. 30a here

For $T \in \mathcal{C}(\tau_0)$, N a process with mean zero and unit variance,

$$\text{MSE}(x) = \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \quad \text{for } x \in \mathbf{R}^d \tag{2.4.3}$$

is the (pointwise) mean square error of $\widehat{T}$ (see also (1.4.2)). We first derive some expressions for the bias, $\mathcal{B}(x)$, and the variance $\mathcal{V}(x)$, of $\widehat{T}(x)$.

~~Fourier transformation interchanges convolution product and pointwise product (see~~ (1.3.8)), and the estimator $\widehat{T}$ of (2.4.2) thus becomes

$$\begin{aligned}
\widehat{T}(x) &= (2\pi)^{-d} \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \int_{\Theta} ([\{\chi(\theta)\tau(\theta)\}_{\mathcal{R}} + \nu_{\mathcal{R}}(\theta)]/\chi(\theta)) e^{-i\langle x,\theta\rangle}\, d\theta \\
&= (2\pi)^{-d} \int_{\Theta} \tau(\theta) e^{-i\langle x,\theta\rangle}\, d\theta + (2\pi)^{-d} \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \int_{\Theta} \nu_{\mathcal{R}}(\theta)\chi(\theta)^{-1} e^{-i\langle x,\theta\rangle}\, d\theta \\
&\cancel{= T_{\Theta}(x) + \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \int_{\mathcal{R}} N(y) k_{\Theta}(x-y)\, dy,}
\end{aligned} \tag{2.4.4}$$

~~where~~

$$
\begin{aligned}
T_\Theta(x) &= (2\pi)^{-d} \int_\Theta \tau(\theta) e^{-i\langle x,\theta\rangle}\, d\theta \\
k_\Theta(x) &= (2\pi)^{-d} \int_{\hat{}} \chi(\theta)^{-1} e^{-i\langle x,\theta\rangle}\, d\theta.
\end{aligned}
$$

Using arguments similar to those given in the derivation of $\widehat{T}(x)$,

the bias of $\widehat{T}(x)$ now becomes

$$
\begin{aligned}
\mathcal{B}(x) &= \mathbf{E}\{\widehat{T}(x)\} - T(x) \\
&= T_\Theta(x) + \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \mathbf{E}\{\int_\mathcal{R} N(y) k_\Theta(x-y)\, dy\} - T(x) \\
&= -T_{\mathbf{R}^d\setminus\Theta}(x) \\
&= -(2\pi)^{-d} \int_{\mathbf{R}^d\setminus\Theta} \tau(\theta) e^{-i\langle x,\theta\rangle}\, d\theta, \qquad (2.4.6)
\end{aligned}
$$

since N is a mean zero process, and

$$
\lim_{\mathcal{R}\uparrow\mathbf{R}^d} \mathbf{E} \int_\mathcal{R} N(y) k_\Theta(x-y)\, dy = \mathbf{E} \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \int_\mathcal{R} N(y) k_\Theta(x-y)\, dy.
$$

Similarly,
For the variance $\mathcal{V}(x)$ of $\widehat{T}(x)$ one obtains

$$
\begin{aligned}
\mathcal{V}(x) &= \operatorname{var} \widehat{T}(x) \\
&= \operatorname{var} \{ \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \widehat{T}_\mathcal{R}(x)\} \\
&= \mathbf{E} \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \{\int_\mathcal{R} N(y) k_\Theta(x-y)\, dy\}^2 \\
&= \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \mathbf{E}\{\int_\mathcal{R} N(y) k_\Theta(x-y)\, dy\}^2, \qquad (2.4.7)
\end{aligned}
$$

from (2.4.2) and (2.4.6). For $x \in \mathbf{R}^d$, the mean square error therefore is

$$
\mathrm{MSE}(x) = (2\pi)^{-2d} \left\{ \int_{\mathbf{R}^d\setminus\Theta} \tau(\theta) e^{-i\langle x,\theta\rangle}\, d\theta \right\}^2 + \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \mathbf{E} \left\{ \int_\mathcal{R} N(y) k_\Theta(x-y)\, dy \right\}^2 . \qquad (2.4.8)
$$

For the convenience of the reader, we summarise our assumptions on images, blur and noise before stating the results.

A1 The envelope τ_0 of the class of images $\mathcal{C}(\tau_0)$,

$$
\mathcal{C}(\tau_0) = \{T \in L^1(\mathbf{R}^d, \mathbf{R}) : |\tau(\theta)| \le \tau_0(\theta) \;\; \forall \theta \in \mathbf{R}^d\},
$$

is in $L^1(\mathbf{R}^d, \mathbf{R}_+)$ and is symmetric.

A2 The point spread function H is a symmetric real-valued function defined on $\mathbf{R}^d$ with $\int H(x)dx = 1$. The Fourier transform of H is χ, and for smoothing sets $\Theta \subset \mathbf{R}^d$ $\chi^{-1}\mathcal{I}_\Theta, k_\Theta \in L^1(\mathbf{R}^d)$.

A3 Either the image $T \in L^2(\mathbf{R}^d)$ for $T \in \mathcal{C}(\tau_0)$, or the point spread function $H \in L^2(\mathbf{R}^d)$.

A4 The noise N is a mean zero, unit variance, (second order) stationary random process with continuous sample paths, a family of autocovariance functions $\gamma_N = \{\gamma_{N,t}\}$ such that $\gamma_N^{(-1)} \in L^1(\mathbf{R}^d)$ and a family of spectral densities $f_N = \{f_{N,t}\} \subseteq L^1(\mathbf{R}^d)$.

A5 The family $f_N = \{f_{N,t}\}$ satisfies

$$|f_{N,t}(\theta)| \leq c_1 t^d \quad \text{for } \theta \in \mathbf{R}^d, c_1 > 0$$

and

$$\inf_{\|\theta\| \leq c_2 t^{-1}} f_{N,t}(\theta) \geq c_3 t^d \quad (c_2, c_3 > 0).$$

Note that **A4** and **A5** incorporate **N1–N5** given earlier.

Proposition 2.2 *Assume that* τ_0, H *and* N *satisfy* **A1**, **A2** *and* **A4**. *Then*

$$\sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\hat{T}(x) - T(x)\}^2 \leq (2\pi)^{-d}\left[\left\{\int_{\mathbf{R}^d\setminus\Theta} \tau_0(\theta)d\theta\right\}^2 + \int_\Theta \chi(\theta)^{-2} f_N(\theta)d\theta\right].$$

The proof of this proposition is given in Subsection 2.6.1. The upper bound for the mean square error $\text{MSE}(x) = \mathbf{E}\{\hat{T}(x) - T(x)\}^2$ can be ~~sharpened~~ expressed in terms of the parameter t, if one also assumes that $f_N = \{f_{N,t}\}$ satisfies **A5**.

Corollary 2.3 *If* τ_0, H, N *satisfy the assumptions of Proposition 2.2. and* $\{f_{N,t}\}$ *satisfies* **A5**, *then there exists a constant* $c_4 > 0$ *such that for* $t > 0$,

$$\sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\hat{T}(x) - T(x)\}^2 \leq c_4\left[\left\{\int_{\mathbf{R}^d\setminus\Theta} \tau_0(\theta)\, d\theta\right\}^2 + t^d \int_\Theta \chi(\theta)^{-2}\, d\theta\right].$$

Note that $0 \leq \mathcal{V}(x) \leq (2\pi)^{-d}\int_\Theta \chi(\theta)^{-2} f_N(\theta)d\theta$ by (2.4.7) and the proof of Proposition 2.2. The upper bound assumption **A5** for f_N, namely

$$|f_{N,t}(\theta)| \leq c_1 t^d,$$

implies now that for $t > 0$, and $\mathcal{V} = \mathcal{V}_t$,

$$\mathcal{V}(x) = |\mathcal{V}(x)| \leq (2\pi)^{-d}\int_\Theta \chi(\theta)^{-2}|f_{N,t}(\theta)|\, d\theta \leq c_4 t^d \int_\Theta \chi(\theta)^{-2}\, d\theta$$

for some constant c_4 depending on c_1. □

To obtain a lower bound for the mean square error, we consider arbitrary estimators $\widetilde{T}$ and regard the estimation problem as one of discriminating between pairs of images T_1 and T_2.

For χ, f_N, τ_0, take $T_1, T_2 \in \mathcal{C}(\tau_0)$, and define

$$T_M = \max\{T_1, T_2\}, \quad \tau_M = FT(T_M); \tag{2.4.9}$$

$$\sigma^2 = (2\pi)^{-d} \int_{\mathbf{R}^d} \tau_M(\theta)^2 \chi(\theta)^2 f_N(\theta)^{-1} \, d\theta. \tag{2.4.10}$$

Using this notation, one obtains the following estimate for the lower bound.

Proposition 2.4 *Assume that τ_0, H, N satisfy* **A1–A4**, *and that N is Gaussian. If σ^2 is given by (2.4.10) and $T_1, T_2 \in \mathcal{C}(\tau_0)$, then for any estimator $\widetilde{T}$*

$$\max_{T=T_1,T_2} \sup_{x \in \mathbf{R}^d} \mathbf{E}_T\{\widetilde{T}(x) - T(x)\}^2 \geq \mathbf{P}\{Z \geq \sigma/2\} T_M(0)^2/8,$$

where Z denotes a standard normal random variable, and $\mathbf{E}_T$ denotes expectation given that the true image is T.

The proof of Proposition 2.4 is given in Subsection 2.6.2.

In Proposition 2.4 and Corollary 2.5 below the lower bound is given in terms of T_M at $x = 0$. A similar result could be proved for arbitrary x. We have chosen $x = 0$, since

$$T_M(0) = (2\pi)^{-d} \int \tau_M(\theta) \, d\theta$$

which is the object of interest when constructing bounds in Theorem 2.6.

Corollary 2.5 *If $\sigma^2 \leq c_5$* as $t \to 0$ *and the assumptions of Propositions 2.4 hold, then there exists a constant $c_6 > 0$ such that*

$$\max_{T=T_1,T_2} \sup_{x \in \mathbf{R}^d} \mathbf{E}_T\{\widetilde{T}(x) - T(x)\}^2 \geq c_6 T_M(0)^2.$$

2.4.2 Optimality of the Partial Fourier Inversion Estimator

In Theorem 2.6 below we show that under suitable regularity conditions the method of partial Fourier inversion restores blurred and noisy images optimally. Here optimality is interpreted in a mean square error minimax sense (for details see Hall (1989)) and may be regarded as the continuous analogue of (1.4.4). Instead of the sample size n, we express the rate of convergence here as a function of the noise parametrisation t as $t \to 0$

(or equivalently as a function of $s = t^{-1}$ as $s \to \infty$). For estimators $\widetilde{T}$ of $T \in \mathcal{C}(\tau_0)$, $\rho : \mathbf{R}_+ \to \mathbf{R}_+$ is an *optimal rate of convergence in a mean square error minimax sense* if there exist constants $c_1, c_2 > 0$ such that, as $s \to \infty$,

1. $\sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}^d} \mathbf{E}\{\widetilde{T}(x) - T(x)\}^2 \leq c_1 \rho(s)$;
2. $\inf_{\widetilde{T}} \sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}^d} \mathbf{E}\{\widetilde{T}(x) - T(x)\}^2 \geq c_2 \rho(s)$;

where the infimum is taken over all possible choices of estimators for T. An estimator $\widetilde{T}$ of T is called *optimal in a mean square error minimax sense* if it satisfies the first of the two conditions above.

Our theorem is given in two parts: the first part describes an upper bound to the mean square error of $\widehat{T}$, and the second part describes a lower bound to the smallest mean square error of $\widetilde{T}$, where the minimum is taken over all possible choices of $\widetilde{T}$. As we shall see in the examples given at the beginning of Section 2.5, upper and lower bound estimates are in fact of the same order, thereby showing that partial Fourier inversion provides an optimal restoration method in many cases of interest.

A summary of assumptions for τ_0, H, and N is given just before Proposition 2.2. In addition to these assumptions, we shall be considering smoothing sets Θ_δ and their complements $\mathsf{C}\Theta_\delta$, $\delta > 0$, of the form

$$\Theta_\delta \equiv \{\theta \in \mathbf{R}^d : \tau_0(\theta)\chi(\theta)^2 > \delta\}$$

$$\mathsf{C}\Theta_\delta \equiv \mathbf{R}^d \setminus \Theta_\delta = \{\theta \in \mathbf{R}^d : \tau_0(\theta)\chi(\theta)^2 \leq \delta\} \tag{2.4.11}$$

as the underlying sets for the partial Fourier inversion estimator $\widehat{T} = \widehat{T}_{\Theta_\delta}$.

Theorem 2.6 *Assume that τ_0, H, N satisfy* **A1–A5**. *Assume that for some $\alpha > 0$ and a decreasing function $\kappa : \mathbf{R}_+ \to \mathbf{R}$, τ_0 satisfies*

$$\delta^{\alpha-1} \int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta \asymp \kappa(\delta) \quad \text{as } \delta \to 0, \tag{2.4.12}$$

where $\mathsf{C}\Theta_\delta$ is given by (2.4.11). Let $k_1 > 0$. For $\widehat{T} = \widehat{T}_{\Theta_\delta}$, Θ_δ as in (2.4.11), choose the smoothing parameter $\delta = \delta(t)$ such that

$$\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)^2 \chi(\theta)^2 \, d\theta = k_1 t^d. \tag{2.4.13}$$

Then there exists a constant $k_2 > 0$ such that for sufficiently small $t > 0$

$$\sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}^d} \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \leq k_2 \{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta\}^2. \tag{2.4.14}$$

If the noise process N is also Gaussian, then there exists a constant $k_3 > 0$ such that for sufficiently small $t > 0$

$$\inf_{\tilde{T}} \sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}^d} \mathbf{E}\{\tilde{T}(x) - T(x)\}^2 + \{\int_{\|\theta\| > c_2 t^{-1}} \tau_0(\theta)\, d\theta\}^2 \geq k_3 \{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta\}^2. \quad (2.4.15)$$

The proof of this theorem is given in Subsection 2.6.3.

2.4.3 Discussion of Results

Optimality of $\widehat{T}_{\Theta_\delta}$. Observe that the upper and lower bounds are not directly comparable, as the lower bound (2.4.15) contains the extra term $\{\int_{\|\theta\| > c_2 t^{-1}} \tau_0(\theta) d\theta\}^2$. However, in most cases of interest, and in particular in Examples 1–3 of Section 2.5, it turns out that

$$\int_{\|\theta\| > c_2 t^{-1}} \tau_0(\theta)\, d\theta = o\left(\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta\right). \quad (2.4.16)$$

This implies that $\widehat{T}_{\Theta_\delta}$ is, in fact, optimal in a minimax sense.

Explicit expressions for Θ_δ and the rate of convergence. If (2.4.16) holds, the estimates (2.4.14) and (2.4.15) show us how to construct the optimal smoothing set Θ_δ for a class of images $\mathcal{C}(\tau_0)$. They furthermore give us the optimal rate of convergence, again uniformly over $\mathcal{C}(\tau_0)$, namely $\{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta) d\theta\}^2$. In the next section, we calculate this rate explicitly for particular examples.

Rate of decrease of τ_0 and χ. Condition (2.4.12) describes a relationship between the image envelope τ_0 and the Fourier kernel χ of the blur. It holds if τ_0 and χ decrease at about the same rate in their respective tails. In Example 4 below, τ_0 decreases much more quickly than χ. A consequence of this is that (2.4.12) fails to hold. In the proof of Theorem 2.6, (2.4.12) is used to show that $\mathcal{V}$, the variance, and $\mathcal{B}$, the bias, satisfy $\mathcal{V} = O(\mathcal{B}^2)$. As we shall see in Example 4, $\mathcal{B}^2 = o(\mathcal{V})$, and the rate of convergence of mean square error cannot therefore be given by $\mathcal{B}^2$.

Dependence of Θ_δ on t. For a fixed envelope τ_0 and blur function H, condition (2.4.13) tells us how the smoothing set Θ_δ varies with the correlation parameter t of the noise process. As t decreases, the smoothing parameter $\delta = \delta(t)$ decreases, and thus the optimal smoothing set Θ_δ becomes larger with decreasing correlation in the noise process.

2.5 Applications

Theorem 2.6 tells us that the optimal rate of convergence of mean square error is $\{\int_{C\Theta_\delta} \tau_0(\theta)\, d\theta\}^2$ for a large class of images and blur functions. In this section we look at some specific image envelopes τ_0 and blur functions H and exhibit the rate of convergence as a function of the noise parameter t.

We first consider some simple examples in which the blur has a rapidly (that is, polynomially or exponentially) decreasing Fourier kernel χ (see paragraph **A** of Subsection 2.2.2). For these examples we calculate the optimal rate as a function of t. For blur functions with compact support (see paragraph **B** of Subsection 2.2.2) we also give a precise description of the optimal smoothing set Θ_δ which excludes suitably chosen neighbourhoods of points for which χ is zero.

2.5.1 Rapidly Decreasing Fourier Kernels

For smoothing sets Θ_δ of the form (2.4.11),

$$\Theta_\delta = \{\theta \in \mathbf{R}^d : \tau_0(\theta)\chi(\theta)^2 > \delta\}, \quad \delta > 0 \tag{2.5.1}$$

and their complements $C\Theta_\delta = \mathbf{R}^d \setminus \Theta_\delta$, put

$$\mathcal{B}_\delta \equiv \int_{C\Theta_\delta} \tau_0(\theta)\, d\theta. \tag{2.5.2}$$

In fact, the right hand side of (2.5.2) represents an estimate of the absolute value of the bias term $\mathcal{B}(x)$ (see Proposition 2.2). We are justified in using this simplification, since we are only interested in the rate of convergence r of the mean square error, and not in the bias term *per se*. Provided the assumptions of Theorem 2.6 are satisfied, the rate of convergence r of the mean square error is then regarded as a function of t and calculated by $r(t) = r(\mathcal{B}^2_{\delta(t)})$.

The examples given below follow the pattern outlined here. For Θ_δ and $\mathcal{B}_\delta$ given by (2.5.1) and (2.5.2) respectively, we show that the assumptions in Theorem 2.6 are satisfied and then calculate the rate of convergence as described in the conclusions of the theorem, that is, we

1. show that there exist $\alpha > 0$ and $\kappa : \mathbf{R}_+ \to \mathbf{R}$, decreasing, such that

$$\delta^{\alpha-1}\mathcal{B}_\delta \asymp \kappa(\delta) \quad \text{as } \delta \to 0; \tag{2.5.3}$$

2. choose $\delta = \delta(t) > 0$ which satisfies

$$\int_{C\Theta_\delta} \tau_0(\theta)^2 \chi(\theta)^2 \, d\theta = k_1 t^d; \tag{2.5.4}$$

3. calculate the rate of convergence r of mean square error as a function of t

$$r(t) = r(\mathcal{B}^2_{\delta(t)}). \tag{2.5.5}$$

For notational simplicity we shall sometimes omit the normalising factor in χ in the examples given below.

Example 1. Let $d \geq 1$, $\tau_0(\theta) = A(1 + ||\theta||)^{-a}$, $\chi(\theta) = (1 + ||\theta||)^{-b}$, where $A > 0$, $a > d$, $b > \frac{1}{2}d$. Note that $a > d$ guarantees that $\tau_0 \in L^1(\mathbf{R}^d)$ and $b > d/2$ that $\chi \in L^2(\mathbf{R}^d)$—see also assumption **A3** in the paragraph preceding Proposition 2.2.

Now for $\delta > 0$,

$$\Theta_\delta = \{\theta : 1 + ||\theta|| < \delta^{-1/(a+2b)}\},$$

and thus

$$\mathcal{B}_\delta \sim \text{ const } \delta^{(a-d)/(a+2b)}.$$

Take $0 < \alpha < (d + 2b)/(a + 2b)$, then $\kappa(s) = s^{-a+(d+2b)/(a+2b)}$ satisfies (2.5.3). Equation (2.5.4) leads to

$$\int_{C\Theta_\delta} (1 + ||\theta||)^{-2(a+b)} \, d\theta \sim c_1 \delta^{\{2(a+b)-d\}/(a+2b)} = c_2 t^d,$$

for c_1, $c_2 > 0$, and thus the rate of convergence of mean square error is

$$r(t) = t^{2d(a-d)/\{2a+2b-d\}}. \tag{2.5.6}$$

Figure 2.1 shows the negative logarithm z of the convergence rate in the case of Example 1: $z = f(a, b)$ where $f(a, b) = 2d(a - b)/\{2(a + b) - d\}$ and $d = 1$. Note that the rate of convergence increases with increasing z. This rate improves with smoother images and less smooth point spread functions.

Example 2. Let $d \geq 1$, $\tau_0(\theta) = A\exp(-a||\theta||^\nu)$, $\chi(\theta) = \exp(-b||\theta||^\nu)$, where $A > 0$, a, $b > 0$, $\nu \geq d$. For $\delta > 0$, the smoothing set Θ_δ is

$$\Theta_\delta = \{\theta : ||\theta||^\nu < (a + 2b)^{-1} \log \delta^{-1}\}.$$

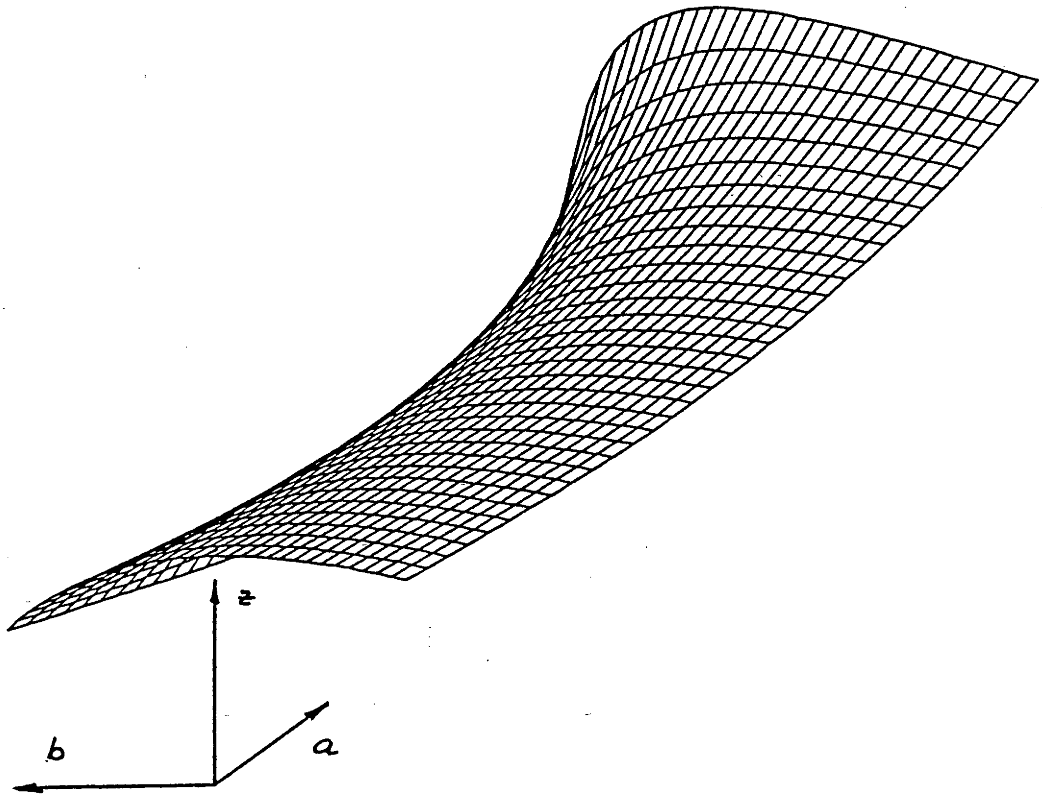

Figure 2.1: Negative logarithm of convergence rate for Example 1.

To obtain an expression for $\mathcal{B}_\delta$, consider

$$\mathcal{B}_\delta \asymp \int_{\mathbb{C}\Theta_\delta} \exp(-a\|\theta\|^\nu)\, d\theta.$$

Put $r = \|\theta\| = \{\Sigma|\theta_i|^2\}^{1/2}$, and consider the change of variable to spherical polar coordinates $\theta = (\theta_1, \ldots, \theta_d)^T \rightsquigarrow (r, \phi_1, \ldots, \phi_{d-1})^T$. One gets

$$d\theta = |J| dr d\phi_1 \ldots d\phi_{d-1},$$

where the absolute value of the Jacobian $|J|$ is proportional to the surface area of the d-dimensional sphere of radius r, that is, $|J| \propto r^{d-1}$.

Writing λ for $(a+2b)^{-1}\log\delta^{-1}$, we obtain

$$\begin{aligned}
\mathcal{B}_\delta &= c_1 \int_{r^\nu > \lambda} \exp(-ar^\nu) r^{d-1} dr \\
&= c_1 \int_{r^\nu > \lambda} (-a\nu)^{-1} r^{-(\nu-1)} r^{d-1}\, d(\exp(-ar^\nu)) \\
&\sim c_2 \lambda^{(d-\nu)/\nu} \exp(-a\lambda) \\
&= c_2 (\log\delta^{-1})^{(d-\nu)/d} \delta^{a/(a+2b)} \qquad \text{for some } c_1, c_2 > 0.
\end{aligned}$$

For $0 < \alpha < 2b/(a+2b)$ and $\kappa(\delta) = (\log\delta^{-1})^{(d-\nu)/\nu} \delta^{\alpha - 2b/(a+2b)}$, one now has

$\delta^{\alpha-1}\mathcal{B}_\delta \asymp \kappa(\delta)$. Furthermore

$$\int_{C\Theta_\delta} \tau_0(\theta)^2\chi(\theta)^2\, d\theta \sim c_3(\log\delta^{-1})^{(d-\nu)/\nu}\delta^{(2a+2b)/(a+2b)} = c_4 t^d$$

(c_3, $c_4 > 0$), which is calculated in a fashion similar to that for the bias term $\mathcal{B}_\delta$. To estimate the rate r of mean square error, we use the fact that

$$r(t) \propto t^{2d}\delta^{-2},$$

to obtain the following rate

$$r(t) = (\log t^{-1})^{(d-\nu)(a+2b)/\nu(a+b)} t^{da/(a+b)}. \tag{2.5.7}$$

In the first two examples we considered envelopes τ_0 and Fourier kernels χ with comparable rates of decrease in their tails. Next we consider two examples in which τ_0 and χ have tails decreasing at different rates.

Example 3. Let $d \geq 1$, $\tau_0(\theta) = A(1+\|\theta\|)^{-a}$, $\chi(\theta) = \exp(-\|\theta\|^b)$, where $A > 0$, $a > d$, $b > 0$. For $\delta > 0$, Θ_δ and $\mathcal{B}_\delta$ are given by

$$\begin{aligned} \Theta_\delta &= \{\theta : \|\theta\|^b < \log\delta^{-1}\} \\ \mathcal{B}_\delta &\sim \text{const } (\log\delta^{-1})^{(d-a)/b}. \end{aligned}$$

Let $0 < \alpha < 1$ and put $\kappa(\delta) = (\log\delta^{-1})^{(d-a)/b}\delta^{\alpha-1}$. Then $\delta^{\alpha-1}\mathcal{B}_\delta \asymp \kappa(\delta)$ as $\delta \to 0$. Methods similar to those used in deriving the bias term $\mathcal{B}_\delta$ in Example 2 above lead to

$$\begin{aligned} \int_{C\Theta_\delta} \tau_0(\theta)^2\chi(\theta)^2\, d\theta &\asymp \int_{\|\theta\|^b \geq \log\delta^{-1}} (1+\|\theta\|)^{-2a}\exp(-2\|\theta\|^b)\, d\theta \\ &\sim c_1(\log\delta^{-1})^{(d-b-2a)/b}\delta^2 \\ &= c_2 t^d \quad (c_1, c_2 > 0). \end{aligned}$$

From this last equation one obtains the rate r of mean square error

$$r(t) = (\log t^{-1})^{(d-b-2a)/b}. \tag{2.5.8}$$

Figure 2.2 shows the negative logarithm z of the convergence rate for Example 3.

Example 4. Let $d \geq 1$, $\tau_0(\theta) = A\exp(-\|\theta\|^a)$, $\chi(\theta) = (1+\|\theta\|)^{-b}$, where $A > 0$, $a > 0$, $b > \frac{1}{2}d$. As before, for $\delta > 0$ one obtains

$$\begin{aligned} \Theta_\delta &= \{\theta : \|\theta\|^a < \log\delta^{-1}\} \\ \mathcal{B}_\delta &\sim \text{const } (\log\delta^{-1})^{(d-a)/a}\delta. \end{aligned}$$

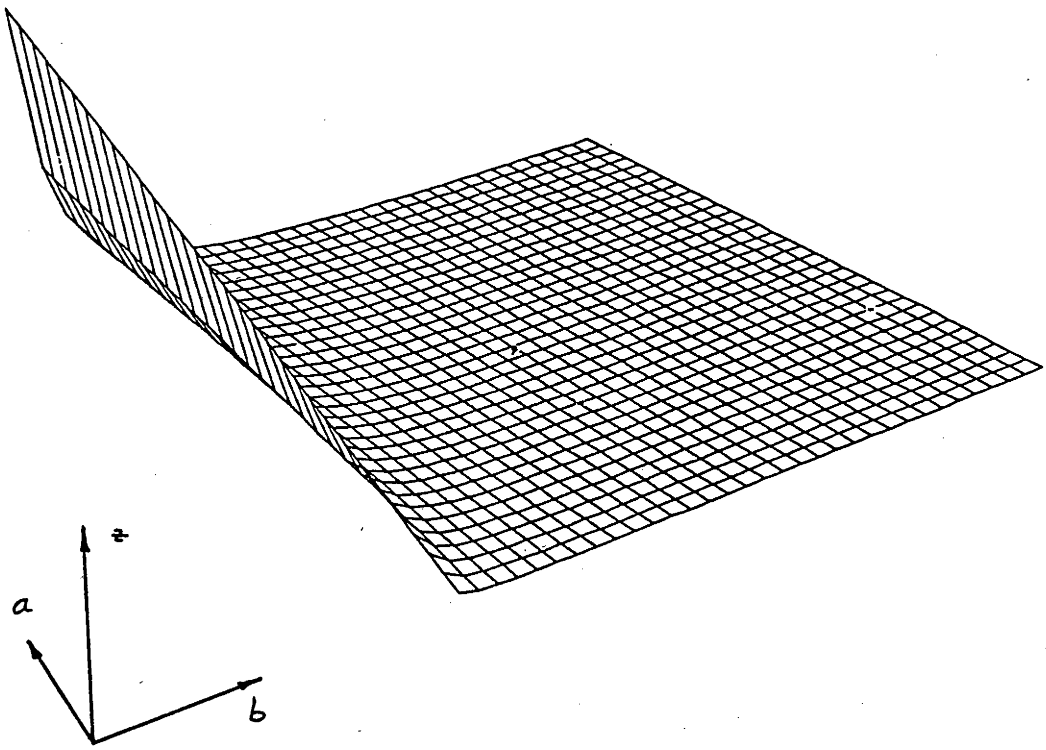

Figure 2.2: Negative logarithm of convergence rate for Example 3.

However, in this case condition (2.5.3) fails: for every $\alpha > 0$, $\delta^{\alpha-1}\mathcal{B}_\delta$ does not tend to infinity as $\delta \to 0$.

Proceeding as in the previous examples, one may show that (2.5.4) leads to

$$(\log \delta^{-1})^{(d-a-2b)/a}\delta^2 = \text{ const } t^d.$$

A calculation of the variance $\mathcal{V}_\delta$, here taken to be $\mathcal{V}_\delta = t^d \int_{\Theta_\delta} \chi(\theta)^{-2} d\theta$, gives

$$\mathcal{V}_\delta \sim \text{ const } t^d (\log \delta^{-1})^{(2b+d)/a}.$$

A comparison of $\mathcal{B}_\delta^2$ and $\mathcal{V}_\delta$ now shows that

$$\mathcal{B}_\delta^2 = o(\mathcal{V}_\delta),$$

and thus the conclusion of Theorem 2.6 cannot apply.

2.5.2 Fourier Kernels with Zeroes

In the previous examples several important point spread functions including the normal, where $\chi(\theta) = \exp(-\|\theta\|^2)$, were considered. In other cases, the Fourier kernel may have zeroes for finite θ. The following example will show the complexity of the smoothing set under these circumstances as well as the dependence of the rate of convergence of mean square error on the smoothness class of the image envelope τ_0 and the Fourier kernel χ. The point spread functions and Fourier kernel considered are described in paragraph **B** of Subsection 2.2.3.

For simplicity let $d = 1$. Assume that τ_0 and H are given by

$$\tau_0(\theta) = A(1+|\theta|)^{-a} \quad \text{for } \theta \in \mathbb{R}, \quad A > 0, \quad a > d = 1; \tag{2.5.9}$$

$$H(x) = H_\nu(x) = \begin{cases} c_1(\nu)\{\cos(\pi x/2)\}^{\nu-1} & \text{for } x \in [-1,+1] \\ 0 & \text{for } x \notin [-1,+1], \end{cases} \tag{2.5.10}$$

where

$$c_1(\nu) = \left[\int_{-1}^{1} \{\cos(\pi x/2)\}^{\nu-1}\, dx\right]^{-1}.$$

The smoothing set Θ is constructed in the following way: for $J > 0$, $0 < \epsilon_j \le 1$, $j \in \mathbb{N}$, $1 \le j \le J$, put

$$\begin{aligned} I_j &= [(\nu-1)\pi/2 + j\pi - \epsilon_j, (\nu-1)\pi/2 + j\pi + \epsilon_j]; \\ \Theta_+ &= [0, J\pi] \setminus \bigcup_{1\le j\le J} I_j, \\ \Theta_- &= \{\theta \in \mathbb{R} : -\theta \in \Theta_+\}; \\ \Theta &= \Theta_+ \cup \Theta_-. \end{aligned} \tag{2.5.11}$$

The optimal smoothing set will be of the form (2.5.11); however the choice of J and ϵ_j will depend on the noise parameter t, ν and the smoothness class a of τ_0 .

Define $r : \mathbb{R}_+ \to \mathbb{R}$ by

$$r(t) = \begin{cases} t^{2(a-1)/(2a+2\nu-1)} & \text{if } 2\nu - a + 2 > 0 \\ t^{2/3} & \text{if } 2\nu - a + 2 < 0 \\ t^{2/3}(\log t^{-1})^{4/3} & \text{if } 2\nu - a + 2 = 0. \end{cases} \tag{2.5.12}$$

For the set-up just described, the theorem below is a special version of Theorem 2.6.

Theorem 2.7 *Assume that τ_0 and H are given by (2.5.9) and (2.5.10), respectively, and that N satisfies the conditions of Theorem 2.6. For $\widehat{T} = \widehat{T}_\Theta$, Θ as in (2.5.11), there* *exists a constant* $k_1 > 0$ *such that*

$$\sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}} \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \leq k_1 r(t) \tag{2.5.13}$$

as $t \to 0$, *where* $r(t)$ *is given by (2.5.12).*

If the noise process is also Gaussian, then there exists a constant $k_2 > 0$ *such that*

$$\inf_{\widetilde{T}} \sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}} \mathbf{E}\{\widetilde{T}(x) - T(x)\}^2 \geq k_2 r(t) \tag{2.5.14}$$

as $t \to 0$.

Figure 2.3 shows the negative logarithm z of the convergence rate for the example of Theorem 2.7. As can be seen, the rate now varies with a and ν in a much more complex way than the rates of the examples treated in the previous subsection.

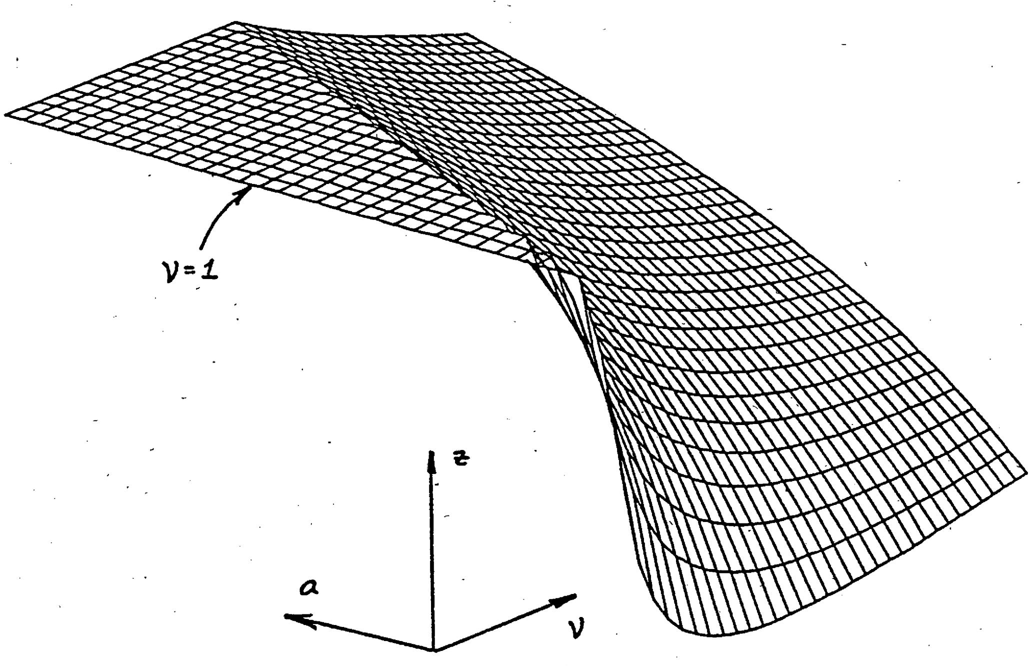

Figure 2.3: Negative logarithm of convergence rate for the example of Theorem 2.7.

The proof of Theorem 2.7 is given in Subsection 2.6.4. In contrast with Theorem 2.6, we have not assumed here that

$$\delta^{\alpha-1} \int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta \asymp \kappa(\delta) \quad \text{as } \delta \to 0,$$

the condition which failed to hold in Example 4 above. As a consequence, we have to calculate rates for $\mathcal{V}$ and $\mathcal{B}$ separately for the three possible cases given in (2.5.12).

2.6 Proofs

This section contains proofs of the propositions and theorems given in the previous two sections. For the convenience of the reader, the results are re-stated before proof.

2.6.1 Proof of Proposition 2.2

Proposition 2.2 *Assume that τ_0, H and N satisfy* **A1**, **A2** *and* **A4**. *Then*

$$\sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\hat{T}(x) - T(x)\}^2 \leq (2\pi)^{-d}\left[\left\{\int_{\mathbf{R}^d\setminus\Theta} \tau_0(\theta)d\theta\right\}^2 + \int_\Theta \chi(\theta)^{-2} f_N(\theta)d\theta\right].$$

Proof of Proposition 2.2

Fix τ_0. For $x \in \mathbf{R}^d$, let $\mathcal{B}(x)$ denote the bias of $\hat{T}$ at x. From (2.4.6) it follows that for $\Theta \subseteq \mathbf{R}^d$

$$\begin{aligned}|\mathcal{B}(x)|^2 &= (2\pi)^{-2d}|\int_{\mathbf{R}^d\setminus\Theta} \tau(\theta)e^{-i\langle x,\theta\rangle}\, d\theta|^2 \\ &\leq (2\pi)^{-2d}\{\int_{\mathbf{R}^d\setminus\Theta} |\tau(\theta)|\, d\theta\}^2 \leq (2\pi)^{-2d}\{\int_{\mathbf{R}^d\setminus\Theta} \tau_0(\theta)\, d\theta\}^2,\end{aligned}$$

since $T \in \mathcal{C}(\tau_0)$. The last expression is independent of $x \in \mathbf{R}^d$ and $T \in \mathcal{C}(\tau_0)$, and thus

$$\sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} |\mathcal{B}(x)|^2 \leq (2\pi)^{-2d}\{\int_{\mathbf{R}^d\setminus\Theta} \tau_0(\theta)\, d\theta\}^2. \tag{2.6.1}$$

To estimate the variance $\mathcal{V}(x)$ of $\hat{T}(x)$, recall from (2.4.7) that

$$\mathcal{V}(x) = \lim_{\mathcal{R}\uparrow\mathbf{R}^d} \mathbf{E}\{\int_{\mathcal{R}} N(y)k(x-y)\, dy\}^2,$$

where $k = k_\Theta$ is given by

$$k(x) = (2\pi)^{-d}\int_\Theta \chi(\theta)^{-1}e^{-i\langle x,\theta\rangle}\, d\theta \quad \text{(see (2.4.5))}.$$

Consider $\mathcal{V}_{\mathcal{R}}(x) \equiv \int_{\mathcal{R}}\int_{\mathcal{R}} \mathbf{E}\{N(x-u)N(x-v)\}k(u)k(v)\, dudv$. Let $f(u,v)$ denote the integrand in $\mathcal{V}_{\mathcal{R}}(x)$ and let $g(u,v) = |f(u,v)|$. By p106 of Dudley (1989), $f \in L^1(\mathbf{R}^d \times \mathbf{R}^d)$. Furthermore, f is dominated (in modulus) by g. The latter is a positive measurable function and therefore Fubini's Theorem (see Theorem 4.4.5 of Dudley (1989)) yields

$$\begin{aligned}\mathbf{E}\int_{\mathcal{R}\times\mathcal{R}} g(u,v)\, dudv &\equiv \int_\Omega\int_{\mathcal{R}\times\mathcal{R}} g(u,v)\, d(u\times v\times\mathbf{P}) \\ &= \int\int g(u,v)\, d\mathbf{P}d(u\times v).\end{aligned}$$

Observe that

$$\int_\Omega g(u,v)\ d\mathbb{P}/leq|k_\Theta(u)k_\Theta(v)|\mathbb{E}N^2 < \infty \qquad a.e.$$

with respect to $du \times dv$. This shows that $f(u,v) \in L^1(\mathbb{R}^d \times \mathbb{R}^d \times \Omega)$, and therefore

$$\begin{aligned}
\mathcal{V}_\mathcal{R} &= \int_\mathcal{R}\int_\mathcal{R} \gamma_N(u-v)k(u)k(v)\ dudv \qquad\qquad \text{(by \textbf{A4})} \\
&= (2\pi)^{-d}\int_\mathcal{R}\int_\mathcal{R}\int_{\mathbb{R}^d} f_N(\theta)k(u)k(v)e^{-i\langle u-v,\theta\rangle}\ dudvd\theta \qquad \text{(by (2.3.3))} \\
&= (2\pi)^{-d}\int_{\mathbb{R}^d} f_N(\theta)\int_\mathcal{R} k(u)e^{-i\langle u,\theta\rangle}\ du\int_\mathcal{R} k(v)e^{i\langle v,\theta\rangle}\ dvd\theta \\
&\to (2\pi)^{-d}\int_{\mathbb{R}^d} f_N(\theta)\{\chi(\theta)^{-1}\mathcal{I}_\Theta\}^2\ d\theta \\
&= (2\pi)^{-d}\int_\Theta f_N(\theta)\chi(\theta)^{-2}\ d\theta.
\end{aligned}$$

since $\int_\mathcal{R} k(u)e^{-i\langle u,\theta\rangle}\ du \to \chi(\theta)^{-1}\mathcal{I}_{Theta}(\theta)$ as $\mathcal{R} \uparrow \mathbb{R}^d$. (See also the derivation of $\widehat{T}$ in Section 2.4.)
Since $\mathcal{V}(x) = \lim_{\mathcal{R}\uparrow\mathbb{R}^d}\mathcal{V}_\mathcal{R}(x)$, it follows that

$$\mathcal{V}(x) = (2\pi)^{-d}\int_\Theta f_N(\theta)\chi(\theta)^{-2}\ d\theta.$$

Thus $\mathcal{V}(x)$ is independent of $x \in \mathbb{R}^d$ and $T \in \mathcal{C}(\tau_0)$. This implies that

$$\sup_{T\in\mathcal{C}(\tau_0)}\sup_{x\in\mathbb{R}^d}\mathcal{V}(x) = (2\pi)^{-d}\int_\Theta f_N(\theta)\chi(\theta)^{-2}\ d\theta. \tag{2.6.2}$$

The desired result now follows from (2.6.1) and (2.6.2). □

2.6.2 Proof of Proposition 2.4

Proposition 2.4 *Assume that τ_0, H, N satisfy* **A1–A4**, *and that N is Gaussian. If σ^2 is given by (2.4.10) and $T_1, T_2 \in \mathcal{C}(\tau_0)$, then for any estimator $\widetilde{T}$*

$$\max_{T=T_1,T_2}\sup_{x\in\mathbb{R}^d}\mathbf{E}_T\{\widetilde{T}(x) - T(x)\}^2 \geq \mathbf{P}\{Z \geq \sigma/2\}T_M(0)^2/8, \tag{2.6.3}$$

where Z denotes a standard normal random variable, and $\mathbf{E}_T$ denotes expectation given that the true image is T.

The following lemma will be used in the proof of this proposition.

Lemma 2.6.1 *Assume the conditions of Proposition 2.4. Let* $\tilde\tau = \tau_2 - \tau_1$, $\tilde B = B_2 - B_1$,

$$\rho^2 = (2\pi)^{-d}\int_{\mathbb{R}^d}|\tilde\tau(\theta)|^2\chi(\theta)^2 f_N(\theta)^{-1}\ d\theta$$

and

$$W = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} N(x)\{\gamma_N(x-y)\}^{(-1)}\widetilde{B}(y)\ dxdy. \qquad \text{lim in mean square}$$

Then with $\gamma_N(x)^{(-1)} = (2\pi)^{-d}\int f_N(\theta)^{-1} e^{-i\langle x,\theta\rangle} d\theta$

1. $\rho^2 = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d} B_2(x)\{\gamma_N(x-y)\}^{-1}B_2(y)\,dxdy$; and

2. W *is a mean zero Gaussian random variable with variance* ρ^2.

Proof of Lemma 2.6.1

To prove part 1, put $b_{\mathcal{R}} = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d}\tilde{B}(x)\{\gamma_N(x-y)\}^{(-1)}\tilde{B}(y)\mathcal{I}_{\mathcal{R}\times\mathcal{R}}(x,y)\,dxdy$. Then

$$b = \lim_{\mathcal{R}\uparrow\mathbf{R}^d} b_{\mathcal{R}} = \int_{\mathbf{R}^d}\int_{\mathbf{R}^d}\tilde{B}(x)\{\gamma_N(x-y)\}^{(-1)}\tilde{B}(y)\,dxdy$$

is well-defined. This can be seen by applying Lebesgue's Dominated Convergence Theorem and Fubini's Theorem with the dominating function $g(x,y) = |\tilde{B}(x)\{\gamma_N(x-y)\}^{(-1)}\tilde{B}(y)| \in L^1(\mathbf{R}^d\times\mathbf{R}^d)$ together with the fact that $\tilde{B}\in L^2(\mathbf{R}^d)$ and $\gamma_N^{(-1)}\in L^1(\mathbf{R}^d)$. It now follows from Parseval's identity (1.3.9) that

$$\begin{aligned} b &= \int_{\mathbf{R}^d}\tilde{B}(x)dx\int_{\mathbf{R}^d}\{\gamma_N(x-y)\}^{(-1)}\tilde{B}(y)\,dy \\ &= \int_{\mathbf{R}^d}\tilde{B}(x)\{\tilde{B}\star\gamma_N^{(-1)}\}(x)\,dx \\ &= (2\pi)^{-d}\int_{\mathbf{R}^d}\tilde{\beta}(\theta)\bar{\tilde{\beta}}(\theta)f_N(\theta)^{-1}\,d\theta, \end{aligned}$$

where $\tilde{\beta}$ denotes the Fourier transform of $\tilde{B}$, and thus $\tilde{\beta} = \chi\tilde{\tau}$, since $\tilde{B} = H\star(T_2-T_1)$. Note that by applying Parseval's identity we made use of the assumption that H or T_i $(i=1,2)$ belongs to $L^2(\mathbf{R}^d)$. It follows that

$$b = (2\pi)^{-d}\int_{\mathbf{R}^d}|\tilde{\tau}(\theta)|^2\chi(\theta)^2 f_N(\theta)^{-1}\,d\theta = \rho^2.$$

For part 2, put $W_{\mathcal{R}} = \int_{\mathcal{R}}\int_{\mathcal{R}} N(x)\{\gamma_N(x-y)\}^{(-1)}\tilde{B}(y)\,dxdy$. Then $W_{\mathcal{R}}$ is Gaussian with $\mathbf{E}W_{\mathcal{R}} = 0$ and finite variance. Similarly to the derivation of $\widehat{T}$ in (2.4.2) as a mean square limit, one shows that $W = \lim W_{\mathcal{R}}$ in mean square as $\mathcal{R}\uparrow\mathbf{R}^d$. Put $\eta^2 = \lim \mathbf{E}W_{\mathcal{R}}^2$. By uniqueness of the limit it follows that

$$\begin{aligned} \eta^2 &= \mathbf{E}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d}\int_{\mathbf{R}^d} N(x)\{\gamma_N(x-y)\}^{-1}\tilde{B}(y)N(s)\{\gamma_N(s-t)\}^{(-1)}\tilde{B}(t)\,dxdydsdt \\ &= \int\int\int\int \gamma_N(x-s)\{\gamma_N(x-y)\}^{-1}\{\gamma_N(s-t)\}^{(-1)}\tilde{B}(y)\tilde{B}(t)\,dxdydsdt, \end{aligned}$$

since $\mathbf{E}\{N(x)N(s)\} = \gamma_N(x-s)$. Thus

$$\begin{aligned} \eta^2 &= \int\tilde{B}(y)\,dy\int\tilde{B}(t)\,dt\int\{\gamma_N(x-y)\}^{(-1)}\,dx\int\gamma_N(x-s)\{\gamma_N(s-t)\}^{(-1)}\,ds \\ &= \int\tilde{B}(y)\,dy\int\tilde{B}(t)\,dt\int\{\gamma_N(x-y)\}^{(-1)}\delta(x-t)\,dx \\ &= \int\int\tilde{B}(y)\tilde{B}(t)\{\gamma_N(y-t)\}^{(-1)}\,dydt \\ &= \rho^2, \end{aligned}$$

by the proof of part 1. □

Proof of Proposition 2.4

It suffices to show (2.6.3) for $x = 0$. Fix $\tau_0 \in L^1(\mathbb{R}^d)$. Take $T_1, T_2 \in \mathcal{C}(\tau_0)$. Without loss of generality we may assume that $T_1 \equiv 0$.

Put $B_i = H \star T_i$, $i = 1,2$. The observations are now of the form $X = B + N$, where B is either the blurred image B_1 or B_2. We want to test two simple hypotheses: $\mathbb{E}X(x) = (H \star T_1)(x)$ and $\mathbb{E}X(x) = (H \star T_2)(x)$. To do this, we first restrict attention to a bounded region $\mathcal{R} \subset \mathbb{R}^d$. Since X is a Gaussian process, we may use the approach of Basawa and Prakasa Rao (1980), p169ff and write the likelihood ratio, say R, as $R = \lim r_n$, where r_n denotes the corresponding likelihood ratio for discrete data. In this case, $r_n = \mu_n/\nu_n$, where μ_n and ν_n now denote likelihoods with respect to Lebesgue measure λ. It follows from Basawa and Prakasa Rao that

$$R = dP/dQ,$$

where P and Q are probability measures on the space of observables with discrete analogues P_n and Q_n respectively such that $\mu_n = dP/d\lambda$ and $\nu_n = dQ_n/d\lambda$.

In the situation considered here, we use the probability measures $L(T_i^{\mathcal{R}})$, $i = 1,2$, given by

$$L(T_i^{\mathcal{R}}) \propto \exp(-Y_i^{\mathcal{R}}/2)$$

where

$$Y_i^{\mathcal{R}} = \int_{\mathcal{R}} \int_{\mathcal{R}} \{X(x) - B_i(x)\}\{\gamma_N(x-y)\}^{(-1)}\{X(y) - B_i(y)\}\, dxdy. \qquad (2.6.4)$$

For a similar set-up see Example 4.1 p177ff of Basawa and Prakasa Rao (1980). Before discriminating between T_1 and T_2, we extend $Y_i^{\mathcal{R}}$ to the region $\mathbb{R}^d$ in the following way: Put

$$\Delta = \lim_{\mathcal{R}\uparrow\mathbb{R}^d} (Y_1^{\mathcal{R}} - Y_2^{\mathcal{R}}), \qquad (2.6.4a)$$

where Δ is regarded as limit in mean square. If the true image is T_1, then

$$\begin{aligned} Y_1^{\mathcal{R}} - Y_2^{\mathcal{R}} &= \int_{\mathcal{R}} \int_{\mathcal{R}} [\{X(x) - B_1(x)\}\{\gamma_N(x-y)\}^{(-1)}\{X(y) - B_1(y)\} \\ &\quad -\{X(x) - B_2(x)\}\{\gamma_N(x-y)\}^{(-1)}\{X(y) - B_2(y)\}]\, dxdy \\ &= \int_{\mathcal{R}} \int_{\mathcal{R}} [2N(x)\{\gamma_N(x-y)\}^{(-1)}(B_2 - B_1)(y) \\ &\quad -(B_2 - B_1)(x)\{\gamma_N(x-y)\}^{(-1)}(B_2 - B_1)(y)]\, dxdy \qquad (2.6.4b) \end{aligned}$$

By part 1 of Lemma 2.6.1,

$$b = \lim_{\mathcal{R}\uparrow\mathbb{R}^d} \int_{\mathcal{R}} \int_{\mathcal{R}} (B_2 - B_1)(x)\{\gamma_N(x-y)\}^{(-1)}(B_2 - B_1)(y)]\, dxdy$$

exists and is well-defined; and the random part in $Y_1^{\mathcal{R}} - Y_2^{\mathcal{R}}$ above has a limit in mean square by part 2 of Lemma 2.6.1. This shows that Δ exists as a limit in mean square.

To test the hypotheses mentioned at the beginning of this proof, by Basawa and Prakasa Rao (1980), Theorem 4.1 and Example 4.1 p169ff, we use the likelihood ratio or Bayes rule, say D_0, for discrimination between T_1 and T_2 in order to decide in favour of T_2 if and only if $\Delta > 0$. The probability p of incorrectly deciding in favour of T_2 given that the true image is T_1 is given by

$$p = \mathbb{P}(\Delta > 0 | T = T_1) \equiv \mathbb{P}_{T_1}(D_0 = 2). \quad (2.6.6)$$

Equations (2.6.4a)(2.6.4b) and (2.6.5) now yield that

$$\begin{aligned} p &= \mathbb{P}[\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 2N(x)\{\gamma_N(x-y)\}^{(-1)}(B_2 - B_1)(y)\, dxdy \\ &\quad \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(B_2 - B_1)(x)\{\gamma_N(x-y)\}^{(-1)}(B_2 - B_1)(y)\, dxdy] \end{aligned}$$

Using the definition of W and ρ^2 given by Lemma 2.6.1 and putting $Z = \rho^{-1}W$,it follows that

$$p = \mathbb{P}\{2W \geq \rho^2\} = \mathbb{P}\{2Z \geq \rho\}.$$

Next, let $\tilde{T}$ denote an estimator of $T = T_1, T_2$. We define a decision rule D for choosing between T_1 and T_2 in the following way:

$$\begin{aligned} D = 1 &\text{ if } |\tilde{T}(0) - T_1(0)| \leq |\tilde{T}(0) - T_2(0)| \\ D = 2 &\text{ if } |\tilde{T}(0) - T_2(0)| \leq |\tilde{T}(0) - T_1(0)|. \end{aligned}$$

If $D = 2$ and $T = T_1$, one has

$$\begin{aligned} |\tilde{T}(0) - T_1(0)| &> |\tilde{T}(0) - T_2(0)| = |T_1(0) - T_2(0) + \tilde{T}(0) - T_1(0)| \\ &\geq |T_1(0) - T_2(0)| - |\tilde{T}(0) - T_1(0)|, \end{aligned}$$

and therefore it follows that

$$|\tilde{T}(0) - T_1(0)| \geq \frac{1}{2}|T_1(0) - T_2(0)| = \frac{1}{2}|\bar{T}(0)|. \quad (2.6.7)$$

Similarly, if $D = 1$ and $T = T_2$, one obtains

$$|\tilde{T}(0) - T_2(0)| \geq \frac{1}{2}|T_1(0) - T_2(0)| = \frac{1}{2}|\bar{T}(0)|, \quad (2.6.8)$$

where $\bar{T} = T_1 - T_2$.

Now let $\mathbf{E}_T$ denote expectation given that the true image is T. Then

$$\begin{aligned} &\max_{T=T_1,T_2} \mathbf{E}_T\{\tilde{T}(0) - T(0)\}^2 \\ &\geq \frac{1}{2}\left[\mathbf{E}_{T_1}\{\tilde{T}(0) - T_1(0)\}^2 + \mathbf{E}_{T_2}\{\tilde{T}(0) - T_2(0)\}^2\right] \\ &\geq \frac{1}{2}\left[\mathbf{E}_{T_1}\{\mathcal{I}_{(D=2)}|\tilde{T}(0) - T_1(0)|^2\} + \mathbf{E}_{T_2}\{\mathcal{I}_{(D=1)}|\tilde{T}(0) - T_2(0)|^2\}\right] \\ &\geq \frac{1}{8}\left[\mathbf{E}_{T_1}\{\mathcal{I}_{(D=2)}\bar{T}(0)^2\} + \mathbf{E}_{T_2}\{\mathcal{I}_{(D=1)}\bar{T}(0)^2\}\right] \qquad \text{(by (2.6.7), (2.6.8)} \\ &\geq \frac{1}{8}\{\mathbb{P}_{T_1}(D=2) + \mathbb{P}_{T_2}(D=1)\}\,\bar{T}(0)^2 \\ &\geq \frac{1}{8}\{\mathbb{P}_{T_1}(D_0=2) + \mathbb{P}_{T_2}(D_0=1)\}\,\bar{T}(0)^2, \end{aligned} \quad (2.6.9)$$

where we have used the Neyman-Pearson lemma for D_0, the Bayes rule (see also Theorem 4.1 of Basawa and Prakasa Rao (1980)) defined in the paragraph preceding (2.6.5), and $\mathcal{I}_{(D=2)}$ (respectively, $\mathcal{I}_{(D=1)}$) denotes the indicator function of the set $D = 2$ (respectively, $D = 1$). Clearly,

$$\mathbf{P}_{T_1}(D_0 = 2) + \mathbf{P}_{T_2}(D_0 = 1) \geq \mathbf{P}_{T_1}(D_0 = 2),$$

and thus combining (2.6.5), (2.6.6) and (2.6.9) yields

$$\begin{aligned} \max_{T=T_1,T_2} \mathbf{E}_T\{\tilde{T}(0) - T(0)\}^2 &\geq \frac{1}{8}\{\mathbf{P}_{T_1}(D_0 = 2)\}\bar{T}(0)^2 \\ &= \frac{1}{8}\mathbf{P}(Z \geq \sigma/2)\bar{T}(0)^2, \end{aligned} \qquad (2.6.9a)$$

To obtain the desired result, take $T_1 \equiv 0$ and let T_2 denote an arbitrary image in $\mathcal{C}(\tau_0)$. This is justified, since we are primarily interested in $T_1(0) - T_2(0)$. By taking $T \equiv 0$, ρ^2 as defined in Lemma 2.6.1 reduces to σ^2 given by

$$\sigma^2 = (2\pi)^{-d} \int_{\mathbb{R}^d} |\tau_2(\theta)|^2 \chi(\theta)^2 f_N(\theta)^{-1} \, d\theta$$

and $\bar{T}(0)$ becomes $\bar{T}(0) = T_2(0)$. From this and (2.6.9a) the desired result follows immediately. □

A similar proof for discrete data and uncorrelated noise is given in Hall (1990).

2.6.3 Proof of Theorem 2.6

Theorem 2.6 *Assume that τ_0, H, N satisfy* **A1–A5**. *Assume that for some $\alpha \geq 0$ and a decreasing function $\kappa : \mathbf{R}_+ \to \mathbf{R}$, τ_0 satisfies*

$$\delta^{\alpha-1} \int_{\complement\Theta_\delta} \tau_0(\theta) \, d\theta \asymp \kappa(\delta) \quad \text{as } \delta \to 0, \qquad (2.6.10)$$

where $\complement\Theta_\delta$ is given by (2.4.11). Let $k_1 > 0$. For $\widehat{T} = \widehat{T}_{\Theta_\delta}$, Θ_δ as in (2.4.11), choose the smoothing parameter $\delta = \delta(t)$ such that

$$\int_{\complement\Theta_\delta} \tau_0(\theta)^2 \chi(\theta)^2 \, d\theta = k_1 t^d. \qquad (2.6.11)$$

Then there exists a constant $k_2 > 0$ such that for sufficiently small $t > 0$

$$\sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \leq k_2\{\int_{\complement\Theta_\delta} \tau_0(\theta) \, d\theta\}^2. \qquad (2.6.12)$$

If the noise process N is also Gaussian, then there exists a constant $k_3 > 0$ such that for sufficiently small $t > 0$

$$\inf_{\tilde{T}} \sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\tilde{T}(x) - T(x)\}^2 + \{\int_{\|\theta\|>c_2 t^{-1}} \tau_0(\theta) \, d\theta\}^2 \geq k_3\{\int_{\complement\Theta_\delta} \tau_0(\theta) \, d\theta\}^2. \quad (2.6.13)$$

Proof of Theorem 2.6

We start with a proof for the lower bound of the mean square error. Recall from Proposition 2.4 and Corollary 2.5 that if

$$\sigma^2 = (2\pi)^{-d} \int_{\mathbf{R}^d} \tau_M(\theta)^2 \chi(\theta)^2 f_N(\theta)^{-1}\, d\theta \le c_5 \tag{2.6.14}$$

for some constant $c_5 > 0$, where $T_M = \max(T_1, T_2)$ with true images $T_1, T_2 \in \mathcal{C}(\tau_0)$, then there exists a constant $c_6 > 0$ such that

$$\max_{T=T_1,T_2} \mathbf{E}\{\tilde{T}(0) - T(0)\}^2 \ge c_6 T(0)^2. \tag{2.6.15}$$

As in the proof of Proposition 2.4 we assume, without loss of generality, that $T_1 \equiv 0$. Now consider $T_2 \in \mathcal{C}(\tau_0)$ such that

$$\tau_2 = \tau_0 \mathcal{I}_\Phi, \tag{2.6.16}$$

where $\Phi \subseteq \mathbf{R}^d$ is symmetric. Then $T = T_M = T_2$ and

$$T(x) = (2\pi)^{-d} \int_{\mathbf{R}^d} \tau_2(\theta) e^{-i\langle x,\theta\rangle}\, d\theta = (2\pi)^{-d} \int_\Phi \tau_0(\theta) e^{-i\langle x,\theta\rangle}\, d\theta.$$

Furthermore, from (2.6.15) and $T = T_2$, it follows that

$$\mathbf{E}\{\tilde{T}(0) - T(0)\}^2 \ge c_7 \{\int_\Phi \tau_0(\theta)\, d\theta\}^2, \tag{2.6.17}$$

where $c_7 = (2\pi)^{-2d} c_6$, provided $\sigma^2 \le c_5$.

Consider

$$L_1 \equiv \sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\tilde{T}(x) - T(x)\}^2 \ge c_7 \sup_{\Phi\in\mathcal{S}_1} \{\int_\Phi \tau_0(\theta)\, d\theta\}^2, \tag{2.6.18}$$

where

$$\mathcal{S}_1 = \{\Phi \subseteq \mathbf{R}^d : \Phi \text{ is symmetric}, \int_\Phi \tau_0(\theta)^2 \chi(\theta)^2 f_N(\theta)^{-1}\, d\theta \le c_5\}.$$

The aim of this part of the proof is to construct a symmetric set $\tilde{\Phi} \in \mathcal{S}_1$ which maximises $\int_\Phi \tau_0(\theta)\, d\theta$ subject to (2.6.11). To achieve this, we construct collections $\mathcal{S}_i$ of symmetric subsets of $\mathbf{R}^d$, the last of which, $\mathcal{S}_4$, will satisfy (2.6.11).

Let B denote the closed d-dimensional L^2-ball of radius $c_2 t^{-1}$, i.e. $B = \{\theta \in \mathbf{R}^d : |\sum \theta_i^2|^{1/2} \le c_2 t^{-1}\}$. By condition **A5** in the list of assumptions prior to Proposition 2.2 in Subsection 2.4.1,

$$f_{N,t}(\theta) \ge c_3 t^d \qquad \text{for } \theta \in B,$$

and it thus follows that

$$L_1 \geq c_7 \sup_{\Phi \in \mathcal{S}_2} \{\int_\Phi \tau_0(\theta)\, d\theta\}^2, \tag{2.6.19}$$

where

$$\mathcal{S}_2 = \{\Phi \subseteq B : \Phi \text{ is symmetric, } \int_\Phi \tau_0(\theta)^2 \chi(\theta)^2\, d\theta \leq c_3 c_5 t^d\}.$$

The inequality (2.6.19) follows from (2.6.18), since $\mathcal{S}_2 \subseteq \mathcal{S}_1$, which can be seen in the following way: for $\Phi \in \mathcal{S}_2$, $f_{N,t}(\theta) \geq c_3 t^d$ whenever $\theta \in \Phi$, and thus for $\Phi \in \mathcal{S}_2$

$$\int_\Phi \tau_0(\theta)^2 \chi(\theta)^2 f_N(\theta)^{-1}\, d\theta \leq c_5,$$

which implies that $\Phi \in \mathcal{S}_1$.

Now consider

$$\mathcal{S}_3 = \{\Phi \subseteq \mathbf{R}^d : \Phi \text{ is symmetric, } \int_\Phi \tau_0(\theta)^2 \chi(\theta)^2\, d\theta \leq c_3 c_5 t^d\},$$

and put $\Phi' = \Phi \cap B$ for $\Phi \in \mathcal{S}_3$. Then

$$\begin{aligned} \sup_{\Phi \in \mathcal{S}_3} \{\int_\Phi \tau_0(\theta)^2\, d\theta\}^2 &\leq 2 \sup_{\Phi \in \mathcal{S}_3} \left[\{\int_{\Phi'} \tau_0(\theta)\, d\theta\}^2 + \{\int_{\Phi \setminus \Phi'} \tau_0(\theta)\, d\theta\}^2 \right] \\ &\leq 2 \sup_{\Phi \in \mathcal{S}_2} \{\int_\Phi \tau_0(\theta)\, d\theta\}^2 + 2\{\int_{\mathbf{R}^d \setminus B} \tau_0(\theta)\, d\theta\}^2, \end{aligned}$$

and hence

$$L \equiv \inf_{\widetilde{T}} L_1 + \{\int_{\mathbf{R}^d \setminus B} \tau_0(\theta)\, d\theta\}^2 \geq c_8 \sup_{\Phi \in \mathcal{S}_3} \{\int_\Phi \tau_0(\theta)\, d\theta\}^2, \tag{2.6.20}$$

where $c_8 = \frac{1}{2}\min(c_7, 1)$.

Choosing $c_5 > 0$ in (2.6.14) large enough such that $c_3 c_5 \geq k_1$, where k_1 is as in (2.6.11), leads to

$$L \geq c_8 \sup_{\Phi \in \mathcal{S}_4} \{\int_\Phi \tau_0(\theta)\, d\theta\}^2, \tag{2.6.21}$$

where

$$\mathcal{S}_4 = \{\Phi \subseteq \mathbf{R}^d : \Phi \text{ is symmetric, } \int_\Phi \tau_0(\theta)^2 \chi(\theta)^2\, d\theta = k_1 t^d\}.$$

Clearly, $\mathcal{S}_4 \subseteq \mathcal{S}_3 \subseteq \mathcal{S}_1$.

Now observe that

$$\sup_{\Phi \in \mathcal{S}_4} \int_\Phi \tau_0(\theta)\, d\theta = \int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta, \tag{2.6.22}$$

where $\mathsf{C}\Theta_\delta = \{\theta \in \mathbf{R}^d : \tau_0(\theta)\chi(\theta)^2 \leq \delta\}$ and $\delta = \delta(t)$ is chosen such that (2.6.11) holds.

To see (2.6.22), one may argue by contradiction: assume there is a set $\Gamma = \{\theta \in \mathbf{R}^d : \tau_0(\theta)\chi(\theta)^2 > \delta\}$ such that (2.6.22) and (2.6.11) hold for Γ instead of $\mathsf{C}\Theta_\delta$. Consider $\theta_1 \in \Gamma$, $\theta_2 \notin \Gamma$, then

$$\tau_0(\theta_1)\chi(\theta_1)^2 > \delta \geq \tau_0(\theta_2)\chi(\theta_2)^2.$$

Let $\mathcal{N}$ denote a neighbourhood of θ_1 of size ϵ^d where ϵ is small enough such that $\mathcal{N} \subseteq \Gamma$. It follows that $\int_{\mathcal{N}} \tau_0$ and $\int_{\mathcal{N}} \tau_0^2\chi^2$ are approximately $\epsilon^d\tau_0(\theta_1)$ and $\epsilon^d\tau_0(\theta_1)^2\chi(\theta_1)^2$, respectively. Let $\mathcal{M}$ denote a neighbourhood of θ_2 of size $\eta^d = \epsilon^d\{\tau_0(\theta_1)\chi(\theta_1)\}^2\{\tau_0(\theta_2)\chi(\theta_2)\}^{-2}$. Note that the size of $\mathcal{M}$ is chosen such that $\int_{\mathcal{M}} \tau_0^2\chi^2$ is approximately $\eta^d\tau_0(\theta_2)^2\chi(\theta_2)^2 = \epsilon^d\tau_0(\theta_1)^2\chi(\theta_1)^2$. (That is, the contribution to the constraint is of the same order.) However, $\int_{\mathcal{M}} \tau_0$ is approximately equal to $\eta^d\tau_0(\theta_2)$, but

$$\eta^d\tau_0(\theta_2) > \epsilon^d\tau_0(\theta_1),$$

which establishes (2.6.22).

This completes the proof of (2.6.13).

In Proposition 2.2 and Corollary 2.3, we showed that an upper bound for MSE(x) is given by

$$U \equiv \sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbf{R}^d} \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \le c_4 \left[\{\int_{\mathbf{R}^d\setminus\Theta} \tau_0(\theta)\, d\theta\}^2 + t^d \int_\Theta \chi(\theta)^{-2}\, d\theta\right],$$

where $\widehat{T} = \widehat{T}_\Theta$ is the partial Fourier inversion estimator, defined on the inversion set $\Theta \subseteq \mathbf{R}^d$. Taking $\Theta = \Theta_\delta = \{\theta \in \mathbf{R}^d : \tau_0(\theta)\chi(\theta)^2 > \delta\}$ as in (2.4.11), $\mathsf{C}\Theta_\delta = \mathbf{R}^d \setminus \Theta_\delta$ therefore gives

$$U \le c_4 \left[\{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta\}^2 + t^d \int_{\Theta_\delta} \chi(\theta)^{-2}\, d\theta\right]. \tag{2.6.23}$$

To obtain the desired result (2.6.12), it thus remains to show that the estimate for the variance is of the same order as that of the square of the bias, that is,

$$t^d \int_{\Theta_\delta} \chi(\theta)^{-2}\, d\theta \le c_9\{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta\}^2 \tag{2.6.24}$$

for some constant $c_9 > 0$.

By the definition of Θ_δ and by (2.6.11), (2.6.24) will follow if we show that

$$\{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)^2\chi(\theta)^2\, d\theta\}\{\int_{\Theta_\delta} \chi(\theta)^{-2}\, d\theta\} \le c_{10}\{\int_{\mathsf{C}\Theta_\delta} \tau_0(\theta)\, d\theta\}^2 \tag{2.6.25}$$

for some constant $c_{10} > 0$, which may depend on k_1 and c_9, and sufficiently small $\delta > 0$.

To show (2.6.24), let Y denote a d-dimensional random vector whose density is proportional to τ_0. Put

$$Z = \tau_0(Y)\chi(Y)^2.$$

Then (2.6.25) is equivalent to

$$\mathbf{E}\{Z\mathcal{I}_{(Z\le\delta)}\}\mathbf{E}\{Z^{-1}\mathcal{I}_{(Z>\delta)}\} \le c_{10}\{\mathbf{P}(Z \le \delta)\}^2. \tag{2.6.26}$$

To show (2.6.26) observe that

$$\begin{aligned}
\mathbf{E}\{Z\mathcal{I}_{(Z\le\delta)}\} &= \int_{z\le\delta} z \, d\mathbf{P}(Z\le z) \le \delta\mathbf{P}(Z\le\delta); \quad \text{and} \\
\mathbf{E}\{Z^{-1}\mathcal{I}_{(Z>\delta)}\} &= \int_{z>\delta} z^{-1} \, d\mathbf{P}(Z\le z) \\
&= -\delta^{-1}\mathbf{P}(Z\le\delta) + \int_{z>\delta} z^{-2}\mathbf{P}(Z\le z)\,dz \\
&\le \int_{z>\delta} z^{-2}\mathbf{P}(Z\le z)\,dz.
\end{aligned}$$

(In the second expectation we used integration by parts.) Applying these estimates to the left hand side of (2.6.26) yields

$$\mathbf{E}\{Z\mathcal{I}_{(Z\le\delta)}\}\mathbf{E}\{Z^{-1}\mathcal{I}_{(Z>\delta)}\} \le \delta\mathbf{P}(Z\le\delta)\int_{z>\delta} z^{-2}\mathbf{P}(Z\le z)\,dz.$$

It remains to show that

$$\delta\int_{z>\delta} z^{-2}\mathbf{P}(Z\le z)\,dz \le c_{10}\mathbf{P}(Z\le\delta). \tag{2.6.27}$$

We now make use of assumption (2.6.10) of the theorem and apply it here to the random vector $Z = \tau_0(Y)\chi(Y)^2$. (Recall that we assumed that the density of Y is proportional to τ_0.) Thus, for some $\alpha > 0$,

$$z^{\alpha-1}\mathbf{P}(Z\le z) \asymp \kappa(z) \qquad \text{as } z\to 0,$$

where $\kappa \nearrow \infty$ as $z \to 0$. Using this relationship, (2.6.27) will follow if we show that

$$\delta\int_{z>\delta} z^{-1-\alpha}\kappa(z)\,dz \le c_{11}\delta^{1-\alpha}\kappa(\delta) \tag{2.6.28}$$

for some constant $c_{11} > 0$. The left hand side of (2.6.28) is estimated by

$$\begin{aligned}
\delta\int_{z>\delta} z^{-1-\alpha}\kappa(z)\,dz &\le \delta\kappa(\delta)\int_{z>\delta} z^{-1-\alpha}\,dz \quad (\text{since } \kappa\nearrow\infty \text{ as } z\to 0) \\
&= \delta^{1-\alpha}\kappa(\delta),
\end{aligned}$$

which shows that (2.6.28) holds. The last calculations imply that (2.6.24) holds, and this together with (2.6.23) completes the proof of the theorem. □

2.6.4 Proof of Theorem 2.7

For the convenience of the reader, we repeat some notation here before stating and proving Theorem 2.7.

Let $d = 1$. Assume that τ_0 and H are given by

$$\tau_0(\theta) = A(1+|\theta|)^{-a} \quad \text{for } \theta\in\mathbf{R} \;\; A>0, \;\; a>d=1, \tag{2.6.29}$$

$$H(x) = H_\nu(x) = \begin{cases} c_1(\nu)\{\cos(\pi x/2)\}^{\nu-1} & \text{for } x \in [-1,+1] \\ 0 & \text{for } x \notin [-1,+1]. \end{cases}, \tag{2.6.30}$$

where

$$c_1(\nu) = \left[\int_{-1}^{1} \{\cos(\pi x/2)\}^{\nu-1}\, dx\right]^{-1}.$$

The smoothing set Θ is constructed in the following way: for $J > 0$, $0 < \epsilon_j \leq 1$, $j \in \mathbb{N}$, $1 \leq j \leq J$, put

$$\begin{aligned} I_j &= [(\nu-1)\pi/2 + j\pi - \epsilon_j, (\nu-1)\pi/2 + j\pi + \epsilon_j]; \\ \Theta_+ &= [0, J\pi] \setminus \bigcup_{1\leq j\leq J} I_j, \\ \Theta_- &= \{\theta \in \mathbb{R} : -\theta \in \Theta_+\}; \\ \Theta &= \Theta_+ \cup \Theta_-. \end{aligned} \tag{2.6.31}$$

The optimal smoothing set will be of the form (2.6.31); however the choice of J and ϵ_j will depend on the noise parameter t, ν and the smoothness class a of τ_0.

Define $r : \mathbb{R}_+ \to \mathbb{R}$ by

$$r(t) = \begin{cases} t^{2(a-1)/(2a+2\nu-1)} & \text{if } 2\nu - a + 2 > 0 \\ t^{2/3} & \text{if } 2\nu - a + 2 < 0 \\ t^{2/3}(\log t^{-1})^{4/3} & \text{if } 2\nu - a + 2 = 0. \end{cases} \tag{2.6.32}$$

Theorem 2.7 *Assume that τ_0 and H are given by (2.6.29) and (2.6.30), respectively, and that N satisfies the conditions of Theorem 2.6. For $\widehat{T} = \widehat{T}_\Theta$, Θ as in (2.6.31), there exists a constant $k_1 > 0$ such that*

$$\sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbb{R}} \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \leq k_1 r(t) \tag{2.6.33}$$

as $t \to 0$, where $r(t)$ is given by (2.6.32).

If the noise process is also Gaussian, then there exists a constant $k_2 > 0$ such that

$$\inf_{\widetilde{T}} \sup_{T\in\mathcal{C}(\tau_0)} \sup_{x\in\mathbb{R}} \mathbf{E}\{\widetilde{T}(x) - T(x)\}^2 \geq k_2 r(t) \tag{2.6.34}$$

as $t \to 0$.

We begin the proof of Theorem 2.7 with some properties of the point spread function H. These are given in

Lemma 2.6.2 *Let $H : \mathbb{R} \to \mathbb{R}$ denote the function*

$$H(x) = \begin{cases} \{\cos(\pi x/2)\}^{\nu-1} & x \in [-1,1] \\ 0 & x \in \mathbb{R} \setminus [-1,1]. \end{cases}$$

Then the Fourier transform χ of H satisfies

1. $\chi(\theta) = 2^{1-\nu}\Gamma(\nu)\pi^{-1}\sin(\theta - \frac{\nu-1}{2}\pi)\Gamma(\frac{\theta}{\pi} - \frac{\nu-1}{2})\{\Gamma(\frac{\theta}{\pi} + \frac{\nu+1}{2})\}^{-1}$, *for* $\theta \in \mathbb{R}$.
2. $\chi(\theta) = 0$, *for* $\theta = \pm\{(\nu-1)\pi/2 + n\pi\ ,\ n \in \mathbb{N},\ n \geq 1\}$.
3. $\chi(\theta) \sim (1 + |\theta|)^{-\nu}\sin(\theta - \frac{\nu-1}{2}\pi)$, *as* $|\theta| \to \infty$.

Proof of Lemma 2.6.2

The Fourier transform χ of H is given by

$$\chi(\theta) = 2^{1-\nu}\Gamma(\nu)\Gamma(\frac{\nu+1}{2} + \frac{\theta}{\pi})\{\Gamma(\frac{\nu+1}{2} - \frac{\theta}{\pi})\}^{-1}.$$

(See Bateman Manuscript Project (1954), 1.6 (27).)

Put $z = \theta/\pi - (\nu-1)/2$. The reflection formula for the gamma-function yields

$$\begin{aligned} \Gamma(\frac{\nu+1}{2} - \frac{\theta}{\pi}) &= \Gamma(1-z) \\ &= \pi\{\Gamma(z)\sin(\pi z)\}^{-1} \\ &= \pi\{\sin(\theta - \frac{\nu-1}{2}\pi)\Gamma(\frac{\theta}{\pi} - \frac{\nu-1}{2})\}^{-1}. \end{aligned}$$

Thus χ can be re-written as

$$\chi(\theta) = 2^{1-\nu}\Gamma(\nu)\pi^{-1}\sin(\theta - \frac{\nu-1}{2}\pi)\Gamma(\frac{\theta}{\pi} - \frac{\nu-1}{2})\{\Gamma(\frac{\theta}{\pi} + \frac{\nu+1}{2})\}^{-1}.$$

Next observe that $\sin(\theta - (\nu-1)\pi/2) = 0$ for $\theta = (\nu-1)\pi/2 + n\pi$, $n = \pm 0, 1, 2, \ldots$. Since Γ has simple poles at the points $\theta = (\nu-1)\pi/2 - n\pi$, $n = 0, 1, 2, \ldots$, and Γ^{-1} has simple zeroes at the points $\theta = -(\nu+1)\pi/2 - n\pi$, $n = 0, 1, 2, \ldots$,. combining the zeroes and poles of the three functions sin, Γ and Γ^{-1}, it can be seen that

$$\chi(\theta) = 0 \quad \text{for } \theta = \pm((\nu-1)\pi/2 - n\pi),\ \ n = 1, 2, 3, \ldots.$$

while $\chi(\theta) \neq 0$ for $\theta = -(\nu-1)\pi/2, -(\nu-1)\pi/2 + \pi, \ldots, (\nu-1)\pi/2 - \pi, \ldots, (\nu-1)\pi/2$, by L'Hospital's rule.

As $|\theta| \to \infty$, $\Gamma(\theta) \sim \sqrt{2\pi}\theta^{\theta-1/2}e^{-\theta}$, which is Stirling's formula. Thus, putting $y = \theta/\pi + 1/2 - \nu/2$,

$$\Gamma(\frac{\theta}{\pi} - \frac{\nu-1}{2})\{\Gamma(\frac{\theta}{\pi} + \frac{\nu+1}{2})\}^{-1} = \Gamma(y)\Gamma(y+\nu)^{-1}$$

$$
\begin{aligned}
&\sim (y^{y-1/2}e^{-y})\{(y+\nu)^{y+\nu-1/2}e^{-(y+\nu)}\}^{-1} \\
&\sim y^{-\nu}e^{\nu} \sim (1+|\theta|)^{-\nu}, \text{ as } |\theta| \to \infty,
\end{aligned}
$$

since ν is finite.

The asymptotic expression for χ therefore becomes

$$
\chi(\theta) \sim \sin(\theta - \frac{\nu-1}{2}\pi)(1+|\theta|)^{-\nu} \quad \text{as } |\theta| \to \infty,
$$

which is the desired result. □

Proof of Theorem 2.7

Since τ_0, H and N satisfy the assumptions of Corollary 2.3, an upper bound U for the mean square error is

$$
U \equiv \sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}} \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \leq c_3 \left[\{\int_{\mathbf{R}\setminus\Theta} \tau_0(\theta)\, d\theta\}^2 + t \int_{\Theta} \chi(\theta)^{-2}\, d\theta \right] \tag{2.6.35}
$$

for some constant $c_3 > 0$, where $\{\int_{\mathbf{R}\setminus\Theta} \tau_0(\theta)\, d\theta\}^2$ bounds the squared bias and $t \int_\Theta \chi(\theta)^{-2}\, d\theta$ bounds the variance of $\widehat{T}$.

Consider the second term on the right hand side of (2.6.35). For fixed $\nu > 0$ and $H = H_\nu$ as in (2.6.33), the Fourier transform χ of H is

$$
\chi(\theta) \sim c_1(\nu) \sin(\theta - \frac{\nu-1}{2}\pi)(1+|\theta|)^{-\nu} \quad \text{for } \theta \in \mathbf{R},
$$

by Lemma 2.6.2. For Θ defined by (2.6.34), one now obtains

$$
\mathcal{V} \leq t \int_\Theta \chi(\theta)^{-2}\, d\theta \leq c_4 t \sum_{j=1}^{J} j^{2\nu} \epsilon_j^{-1}, \tag{2.6.36}
$$

since, by Lemma 2.6.2, $\sin(\theta - (\nu-1)\pi/2) = 0$ for $\theta \in \mathcal{N}_\nu = \{\theta : \theta = \pm(\nu-1)\pi/2 + n\pi), n \geq 1\}$, and thus $|\sin(\theta - (\nu-1)\pi/2)| \geq \epsilon_j$ for some $j = 1, \ldots, J$ and $\theta \in \Theta$.

Write $\mathcal{B}$ for the absolute value of the bias. Then

$$
\begin{aligned}
\mathcal{B} &\leq (2\pi)^{-d} \int_{\mathbf{R}\setminus\Theta} \tau_0(\theta)\, d\theta \\
&= (2\pi)^{-d} A \int_{\mathbf{R}\setminus\Theta} (1+|\theta|)^{-a}\, d\theta \\
&\leq c_5 \{\sum_{j=1}^{J} j^{-a}\epsilon_j + J^{-(a-1)}\}
\end{aligned} \tag{2.6.37}
$$

for some $c_5 > 0$, and thus an estimate for U as defined in (2.6.35) is

$$U \le c_6 \left[\left\{ \sum_{j=1}^{J} j^{-a} \epsilon_j + \mathcal{S}^{-(a-1)} \right\}^2 + t \sum_{j=1}^{J} j^{2\nu} \epsilon_j^{-1} \right], \tag{2.6.38}$$

where $c_6 > 0$.

To exhibit a lower bound for the mean square error, we proceed as in the proof of Theorem 2.6 by putting $\tau = \tau_0 \mathcal{I}_{\complement\Theta_\delta}$ as in (2.6.16) and maximising

$$\int_{\complement\Theta} \tau_0(\theta)\, d\theta \text{ subject to } \int_{\complement\Theta} \tau_0(\theta)^2 \chi(\theta)^2\, d\theta = c_7 t.$$

For Θ as in (2.6.31), take

$$\Phi \equiv \complement\Theta \cap [-t^{-1}, t^{-1}], \quad \tau \equiv \tau_0 \mathcal{I}_\Phi, \tag{2.6.39}$$

that is, $\tau(\theta) = A(1+\theta)^{-1}$ for θ (or $-\theta$) $\in I_j$, $1 \le j \le t^{-1}$, and $\tau(\theta) = 0$ otherwise. As in the proof of Theorem 2.6, we deduce that the lower bound L of the mean square error satisfies

$$L \equiv \inf_{\widetilde{T}} \sup_{T \in \mathcal{C}(\tau_0)} \sup_{x \in \mathbf{R}} \mathbf{E}\{\widetilde{T}(x) - T(x)\}^2 \ge c_8 \sup_{\Phi} \{ \int_\Phi \tau_0(\theta)\, d\theta \}^2 \tag{2.6.40}$$

(see (2.6.19)–(2.6.22)) for some $c_8 > 0$. For Φ and τ as in (2.6.39), write $\mathcal{B}_\Phi$ for the absolute value of the bias of τ. Then

$$\mathcal{B}_\Phi = \int_\Phi \tau_0(\theta)\, d\theta \asymp \sum_{j=1}^{t^{-1}} j^{-a} \epsilon_j;$$

$$\int_\Phi \tau_0(\theta)^2 \chi(\theta)^2\, d\theta \asymp \sum_{j=1}^{t^{-1}} j^{-2\nu-2a} \epsilon_j^3;$$

and thus

$$L \ge c_9 \sup_{J, \epsilon_j} \left\{ \left(\sum_{j=1}^{t^{-1}} j^{-a} \epsilon_j \right)^2 \text{ subject to } \sum_{j=1}^{t^{-1}} j^{-2a-2\nu} \epsilon_j^3 \asymp t \right\} \quad (c_9 > 0). \tag{2.6.41}$$

To obtain the optimal rate $r(t)$ postulated in the theorem, we consider three separate cases. For each we choose values for J and ϵ_j (see (2.6.31)) in order to determine an optimal smoothing set Θ. For these values of J and ϵ_j, we calculate rates of convergence for $\mathcal{V}$ and $\mathcal{B}$ and show that $\mathcal{V} = O(\mathcal{B}^2)$. This then implies that $U = O(\mathcal{B}^2)$. For the lower bounds it suffices to calculate $\mathcal{B}$ subject to the constraint $\sum_{j=1}^{t-1} j^{-2a-2\nu} \epsilon_j^3 \asymp t$ as in the proof of Theorem 2.6.

Case I: $2\nu - a + 2 > 0$ (i.e. $a - 1 < 2\nu + 1$). We choose values of α and β in the following way: Let $\alpha > 0$ be such that

$$a - 1 < \alpha(2a + 2\nu - 1) < 2\nu + 1.$$

Put

$$\beta = \alpha(2a + 2\nu - 1), \quad \epsilon_j = \min(1, t^\alpha j^\beta), \quad J = t^{-1/(2a+2\nu-1)}.$$

Using these values of ϵ_j and J in the definition of Θ, one obtains

$$\begin{aligned}
\mathcal{B} &\asymp \sum_{j=1}^{J} j^{-a}\epsilon_j = t^\alpha \sum_{j=1}^{J} j^{-a+\beta} \asymp t^\alpha J^{\beta-a+1} = t^{(a-1)/(2a+2\nu-1)}; \\
\mathcal{V} &\asymp t\sum_{j=1}^{J} j^{2\nu}\epsilon_j^{-1} = t^{1-\alpha}\sum_{j=1}^{J} j^{2\nu-\beta} \asymp t^{1-\alpha}J^{2\nu-\beta+1} = t^{2(a-1)/(2a+2\nu-1)}.
\end{aligned}$$

Hence, $\mathcal{V} = O(\mathcal{B}^2)$ and $U = O(t^{2(a-1)/(2a+2\nu-1)})$.

For the lower bound, put $m = t^{-1/(2a+2\nu-1)}$, and

$$\epsilon_j = \begin{cases} 1 & m \le j \le t^{-1} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$t^{-1}\sum_{j=m}^{t^{-1}} j^{-2a-2\nu}\epsilon_j^3 \asymp t^{-1}m^{-(2a+2\nu-1)} = 1;$$

$$\mathcal{B}_\Phi \asymp \sum_{j=m}^{t^{-1}} j^{-a}\epsilon_j \asymp m^{-a+1} = t^{(a-1)/(2a+2\nu-1)},$$

for Φ as in (2.6.39); and therefore it follows that MSE $\asymp t^{2(a-1)/(2a+2\nu-1)}$.

Case II: $2\nu - a + 2 < 0$ (i.e. $(2\nu - a)/2 < -1$). Put

$$\alpha = 1/3, \quad \beta = (2\nu + a)/2$$

$$\epsilon_j = \min(1, t^\alpha j^\beta), \quad J = t^{-\alpha/\beta}.$$

For the upper bound one obtains

$$\begin{aligned}
\mathcal{B} &\asymp \sum_{j=1}^{J} j^{-a}\epsilon_j = t^\alpha \sum_{j=1}^{J} j^{(2\nu-a)/2} \asymp t^\alpha = t^{1/3}; \\
\mathcal{V} &\asymp t\sum_{j=1}^{J} j^{2\nu}\epsilon_j^{-1} = t^{1-\alpha}\sum_{j=1}^{J} j^{2\nu-\beta} = t^{1-\alpha}\sum_{j=1}^{J} j^{(2\nu-a)/2} \asymp t^{1-\alpha} = t^{2/3},
\end{aligned}$$

and therefore $U = O(t^{2/3})$.

For the lower bound, put $m = t^{-2/\{3(2\nu+a)\}}$, and

$$\epsilon_j = \begin{cases} t^{1/3} j^{(2\nu+a)/2} & \text{for } 1 \le j \le m \\ 0 & \text{otherwise.} \end{cases}$$

Then for $0 \le \epsilon_j \le 1$,

$$t^{-1} \sum_{j=1}^{t^{-1}} j^{-2a-2\nu} \epsilon_j^3 = \sum_{j=1}^{m} j^{(2\nu-a)/2} \asymp 1;$$

$$\mathcal{B}_\Phi \asymp \sum_{j=1}^{m} j^{-a} \epsilon_j = t^{1/3} \sum_{j=1}^{m} j^{(2\nu-a)/2} \asymp t^{1/3},$$

which implies that MSE $\asymp t^{2/3}$.

Case III: $2\nu - a + 2 = 0$ (i.e. $2(\nu+1) = a - 1$). Put

$$\epsilon_j = \min\{1, (t/\log t^{-1})^{1/3} j^{a-1}\}, \quad J = t^{-1}.$$

For the upper bound one obtains

$$\begin{aligned} \mathcal{B} &\asymp \sum_{j=1}^{J} j^{-a} \epsilon_j = (t/\log t^{-1})^{1/3} \sum_{j=1}^{J} j^{-1} = t^{1/3} (\log t^{-1})^{2/3}; \\ \mathcal{V} &\asymp t \sum_{j=1}^{J} j^{2\nu} \epsilon_j^{-1} = t^{2/3} (\log t^{-1})^{1/3} \sum_{j=1}^{J} j^{-1} = t^{2/3} (\log t^{-1})^{4/3}, \end{aligned}$$

and therefore

$$U = O(t^{2/3} (\log t^{-1})^{4/3}).$$

For the lower bound, put $m = t^{-1}$ and

$$\epsilon_j = \begin{cases} (t/\log t^{-1})^{1/3} j^{(2\nu+a)/2} & \text{for } 1 \le j \le m \\ 0 & \text{otherwise.} \end{cases}$$

For $0 \le \epsilon_j \le 1$, we have

$$t^{-1} \sum_{j=1}^{m} j^{-2a-2\nu} \epsilon_j^3 = t^{-1} (t/\log t^{-1}) \sum_{j=1}^{m} j^{-1} \asymp 1;$$

$$\mathcal{B}_\Phi \asymp \sum_{j=1}^{m} j^{-a} \epsilon_j = t^{1/3} (\log t^{-1})^{-1/3} \sum_{j=1}^{m} j^{-1} \asymp t^{1/3} (\log t^{-1})^{2/3},$$

and therefore we conclude that MSE $\asymp t^{2/3} (\log t^{-1})^{4/3}$, as required. $\square$

Chapter 3

Cross-Validation

3.1 Introduction

A common problem of image enhancement, nonparametric regression and density estimation is the selection of a smoothing parameter. Sometimes the smoothing parameter is selected by eye or by using additional information that may be available. Often, however, it is necessary to rely on an automatic, data-based and objective way of choosing the smoothing parameter. In nonparametric regression and density estimation cross-validation has become a well-established and mathematically justified method for choosing the optimal amount of smoothing, and many other methods of determining the smoothing parameter are also now available (see e.g. Golub *et al.* (1979), Stone (1984), Silverman (1985), Silverman (1986), Härdle (1989) and Wahba (1990)). The basic ideas of least squares cross-validation in nonparametric regression have been briefly described in Section 1.4 of Chapter 1.

In the processing of indirect images, automatic methods for choosing the smoothing parameter are still rather scarce. Attempts at obtaining the right amount of smoothing can be found in specific practical problems (see Koch and Tarlowski (1987)). More closely related to the problem discussed in this chapter is a paper by Thompson *et al.* (1990) on the method of generalised cross-validation (GCV) for blurred and noisy data (see also Golub *et al.* (1979)). ~~In the research presented here we are concerned with cross-validation only and not with generalised cross-validation, and consider extensions of cross-validation to image analysis primarily from a theoretical point of view.~~
Insert from opposite page

One of the differences between nonparametric regression and image enhancement is the presence of blur. For this reason, the cross-validation formalism outlined in Chapter 1 cannot directly be applied to image processing problems. Amongst the possible ways of generalising cross-validation, we consider two distinct approaches, both of which are shown to yield asymptotically optimal image enhancement provided the amount of blur in the observed image does not exceed a certain level.

Because of the mathematical complexity of cross-validation in image processing—accommodating the blur in the image adds a new degree of difficulty beyond those already present in nonparametric regression—a very general feasibility study of cross-validation is rather forbidding. This has forced us to make a number of simplifying assumptions in order to arrive at a more tractable analysis:

- the true image is modelled by a smooth deterministic function;
- the point spread function is assumed to be known and has a specific form;
- the noise in the observations is modelled by additive white noise.

Real images are much more complex than smooth deterministic functions. Describing true images in terms of the smoothness class they belong to, however, will allow us to assess the effect of image smoothness on the success of the procedure in a precise manner. Similar conclusions hold for the restriction to a specific class of point spread functions. Under the above assumptions we are able to describe the two generalisations of cross-validation precisely and can give specific bounds on the admissible amount of blur.

We describe our models for the true images and the point spread functions as well as our two methods of cross-validation in Section 3.2. The models will be seen to be closely related to some of those discussed in the previous chapter. In Section 3.3 we present our results on the performance of cross-validation. In Proposition 3.4 we first show that sum of squared error (SSE) and mean sum of squared error (MSSE) are asymptotically the same. Theorems 3.5 and 3.6 give precise conditions under which our two cross-validation methods are, to first order, asymptiotically the same as SSE. As we shall see in these two theorems, the first (naive) cross-validation method performs better than the second method. For the former we then strengthen the mean square results of Theorem 3.5 to *a.s.* results. This is done in Theorems 3.8 and 3.10. We conclude our section of results with some simulations of the two cross-validation methods which confirm our asymptotic results. The proofs of our propositions and theorems are given in Section 3.4, and the Appendix (Section 3.5) contains results on martingales adapted to our needs.

A much condensed version of the mean square results presented here can be found in Hall and Koch (1991).

3.2 Mathematical Models and Cross-Validation in Image Analysis

3.2.1 Mathematical Models for the Observations

Cross-validation is a practical technique which is applied to real data. In this chapter we therefore restrict attention to discretely defined observations. As we shall see below, our images and point spread functions can be regarded as discretisations of the continuously defined models of the previous chapter.

A. The true image t. We assume that the true image is the deterministic function $t : \mathbb{Z}^d \to \mathbb{R}$ whose (discrete) Fourier transform τ is

$$\tau(\theta) \asymp n^d(1 + \|n\theta\|)^{-a} \quad \text{for } \theta \in \Omega \equiv [-\pi, \pi]^d, \quad a > 0. \tag{3.2.1}$$

From (1.3.7) it follows that t is of the form

$$t(j) = (2\pi)^{-d} \int_\Omega \tau(\theta) e^{-i\langle j,\theta\rangle} \, d\theta \quad \text{for } j \in \mathbb{Z}^d.$$

The image t may be interpreted as a digitised version of a continuously defined image $T : \mathbb{R}^d \to \mathbb{R}$ restricted to $J^d = [-1, 1]^d$: for fixed $n > 0$, let G_n denote a regular square based grid on J^d consisting of n^d equally spaced points. For $l \in G_n$ define $\tilde{t}$ on G_n by

$$\tilde{t}(l) = T(l) \quad \text{for } l = (l_1, \ldots, l_d)^T, \quad |l_i| \leq 1, \quad i = 1, \ldots, d, \tag{3.2.2}$$

and put

$$t(j) = \tilde{t}(n^{-1}j) \quad \text{for } n^{-1}j \in G_n. \tag{3.2.3}$$

Equation (3.2.3) defines a function t on a subgrid $\mathcal{G}_n \subseteq \mathbb{Z}^d$ where

$$\mathcal{G}_n = \{j \in \mathbb{Z}^d \ : \ |j_i| \leq n, \ i = 1, \ldots, d\}.$$

Furthermore, the Fourier transforms τ and $\tilde{\tau}$ of t and $\tilde{t}$ respectively are related by

$$\tau(\theta) = n^d \tilde{\tau}(n\theta) \quad \text{for } \theta \in \Omega.$$

In this sense t may be regarded as the digitised version of an a times continuously differentiable image T (see (2.2.2) and paragraph A of Subsection 2.2.1). Furthermore, if $a \geq 1$, then T' is bounded, and T satisfies a *Lipschitz condition*, that is, there exists $M > 0$ such that

$$|T(x) - T(y)| \leq M|x - y| \quad \text{for } x, y \in J^d. \tag{3.2.4}$$

If t is regarded as a discretisation of T onto $\mathcal{G}_n$, then (3.2.4) implies that

$$|t(j) - t(k)| = O(n^{-1}) \text{ for } j, k \in \mathcal{G}_n, \ \|j - k\|_\infty \leq 1. \tag{3.2.5}$$

We shall meet with a condition of this form in Subsection 3.3.3 in the context of the *a.s.* properties of the first cross-validation method.

Varying n in the definition (3.2.1) corresponds to a variation of the pixel grid G_n and increasing the value of n may thus be regarded as increasing the sample size. Furthermore, large values for n imply that t resolves fine details of the image T.

B. The point spread function h. For $0 < \rho < 1$ define a family of point spread functions h by

$$h(j) = \{(1-\rho)/(1+\rho)\}^d \rho^{|j|} \text{ for } j \in \mathbb{Z}^d, \tag{3.2.6}$$

where the parameter ρ is known. The factor $\{(1-\rho)/(1+\rho)\}^d$ ensures that h is a density, and thus h preserves average image intensity. For $\rho > 0$, h may be regarded as the discretisation of the out-of-focus blur H given by

$$H(x) = c_\nu \exp(-\nu|x|) \text{ for } x \in \mathbf{R}^d, \ \ c_\nu > 0, \tag{3.2.7}$$

which is a special case of the point spread functions described in paragraph B of Subsection 2.2.2. Using the above interpretation of h as a discrete version of H implies that $\rho = \exp(-\Delta\nu)$ where Δ denotes the separation between adjacent pixels. Insert from opposite page

Although (3.2.6) is a natural digitisation of the out-of-focus blur (3.2.7), from a mathematical point of view it will be more convenient to reparametrise the family of point spread functions h by putting

$$\lambda = (1-\rho)^{-1} \text{ for } 0 < \rho < 1.$$

This leads to the family $h = \{h_\lambda\}$ of point spread functions

$$h_\lambda(j) = (2\lambda - 1)^{-d}(1 - \lambda^{-1})^{|j|} \text{ for } j \in \mathbb{Z}^d, \ \ \lambda \geq 1, \tag{3.2.8}$$

which we shall use in the sequel. Let χ_λ denote the (discrete) Fourier transform of h_λ, then

$$\chi_\lambda(\theta) = \prod_{i=1}^{d} \{1 + 2\lambda^2(1-\lambda^{-1})(1-\cos\theta_i)\}^{-1} \text{ for } \theta = (\theta_1, \ldots, \theta_d)^T \in \Omega. \tag{3.2.9}$$

We shall often drop the subscript λ and write χ for the Fourier transform of h. As can be seen from (3.2.8), $\lambda = 1$ corresponds to the no-blur case, that is, $h(j) = 1$ for $j = 0$ and $h \equiv 0$ for $j \neq 0$.

C. The observed data X. As in the set-up of the previous chapter, the true image is degraded linearly by the point spread function h to yield the blurred image

$$b(j) = (h \star t)(j) = \sum_{l \in \mathbb{Z}^d} h(l)t(j-l) \quad \text{for } j \in \mathbb{Z}^d. \tag{3.2.10}$$

If $b = h \star t$ is regarded as a discretisation of the blurred scene $B = H \star T$, H and T as in the previous paragraphs, the following effects may occur: if $\nu > 0$ in (3.2.7) is fixed, then this corresponds to assuming that $\lambda(n) \sim const.n$ as the discretisation becomes finer and so $\lambda \to \infty$ as $n \to \infty$. On the other hand, assuming that λ grows at a slower rate than n is equivalent to modelling the effect of digitising the blurred scene B when the amount of blur decreases with increasing sample size.

The blurred image (3.2.10) is further corrupted by random noise ϵ. We assume that the observed data are of the form

$$X_j = (h \star t)(j) + \epsilon_j \quad \text{for } j \in \mathbb{Z}^d, \tag{3.2.11}$$

where the ϵ_j are independent and identically distributed with mean zero and finite variance σ^2.

3.2.2 The Estimator $\hat{t}$

For the discretely defined observations X_j of (3.2.11), the partial Fourier inversion estimator $\hat{t}$ is given by

$$\hat{t}(j) = (2\pi)^{-d} \int_{\Theta} \{\xi(\theta)/\chi(\theta)\} e^{-i\langle j,\theta\rangle}\, d\theta \quad \text{for } j \in \mathbb{Z}^d, \tag{3.2.12}$$

where ξ and χ now denote the discrete Fourier transforms of X and h respectively, and $\Theta = [-\delta,\delta]^d$ denotes an inversion or smoothing set (see (1.3.12)), which is now a subset of $\Omega = [-\pi,\pi]^d$ and depends on the smoothing parameter δ.

Fix $n > 0$, and put

$$\mathcal{R}_n = \{j \in \mathbb{Z}^d \; : \; j = (j_1, \ldots, j_d)^T, \; |j_i| \leq Kn, \; i = 1, \ldots, d\}, \tag{3.2.13}$$

where $K > 0$ is a constant. Fixing K in the definition of $\mathcal{R}_n$ corresponds to considering a fixed region $K \times K \times \ldots \times K$ (d times) in $\mathbf{R}^d$ of the corresponding continuously defined image T. For mathematical simplicity it is sometimes convenient to take $K = 1$, but the analysis is not affected by this.

In this chapter we are interested in assessing global performance of $\hat{t}$ on $\mathcal{R}_n$. In analogy with nonparametric regression we consider the distance measures SSE, sum of squared error, and MSSE, mean sum of squared error, given by

$$\begin{aligned}\mathrm{SSE}(n) &= \sum_{j\in\mathcal{R}_n}\{\hat{t}(j)-t(j)\}^2,\\ \mathrm{MSSE}(n) &= \mathbf{E}\{\mathrm{SSE}(n)\}.\end{aligned} \tag{3.2.14}$$

The variable n indicates restriction to the region $\mathcal{R}_n$, and thus sums of $O(n^d)$ points in $\mathbb{Z}^d$ are considered. If no ambiguity exists, we shall often drop the parameter n, and write $\mathcal{R}$, SSE and MSSE.

For us, to obtain a 'good' estimator $\hat{t}$ of t amounts to choosing the smoothing parameter δ such that $\hat{t}$ is close to t on $\mathcal{R}$ with respect to SSE or MSSE. Both these measures are based on the unknown image t, and they cannot therefore be used in the selection of the smoothing parameter in practice. In the next subsection we describe two methods of selecting a smoothing parameter which are based on the data, and do not rely on knowledge of the true image t.

In the definition (3.2.14), SSE and MSSE are functions of n, which is classically interpreted as the sample size. Sometimes in our analysis, we shall consider sums over $\mathbb{Z}^d$. To distinguish between the finite sums over $\mathcal{R}_n$ and the infinite sums, throughout this chapter we shall always write sums over $\mathcal{R}_n$ in the form $\sum_{j\in\mathcal{R}_n}$, while sums over $\mathbb{Z}^d$ will usually be abreviated in the following way: $\sum_j \equiv \sum_{j\in\mathbb{Z}^d}$. So when the range of the summation variable is omitted, $\mathbb{Z}^d$ is implied.

3.2.3 Two Approaches to Cross-Validation

We now consider two ways of generalising cross-validation to image analysis. These may be regarded as the two extreme cases. The first (naive) method mimics the situation of nonparametric regression in the sense that it ignores the existence of the blur. In contrast to this, the second method takes careful account of the blur by removing the entire blur before the cross-validation estimate is constructed.

Many intermediate ways may be possible; however, we shall only be concerned with the two above-mentioned approaches here for the following reasons. Comparing the results of ignoring the blur completely and of taking careful account of the blur will show clearly how the blur in the image affects cross-validation. In both cases, cross-validation fails for large amounts of blur. In the first method the failure is due to the increase in the bias as the amount of blur is increased, while in the second method cross-validation fails because the removal of blur from the data increases the variance term intolerably. Exactly what 'large' amounts of blur means can be seen in Theorems 3.4 and 3.5. Any other method which takes some account of the blur will not demonstrate the effect of the blur as well as either of the methods mentioned above. This feature together with the very high mathematical complexity that cross-validation in image analysis presents has been the driving force in choosing our particular methods of generalising cross-validation.

For the remainder of this section we restrict attention to the distance measure SSE of (3.2.14).

As indicated in Section 1.4, in cross-validation the data are split into two disjoint parts, and one part is used to assess the performance of an estimator based on the other part. To see how this applies in image analysis, assume that one wants to assess the size of $\hat{t}(j) - t(j)$. If $t(j)$ is estimated by $\tilde{t}(j)$, calculated from a small number of data points in the neighbourhood of j (e.g. $\tilde{t}(j) = X(j)$, as in nonparametric regression), and $\hat{t}^*(j)$, the estimate of $\hat{t}(j)$, is calculated from the complementary part of the data, then $\hat{t}^*(j) - \tilde{t}(j)$ should be close to $\hat{t}(j) - t(j)$. For fixed $n > 0$, this suggests the use of the distance measure

$$\widehat{\mathrm{SSE}} = \sum_{j \in \mathcal{R}} \{\hat{t}^*(j) - \tilde{t}(j)\}^2 \tag{3.2.15}$$

as an approximation to SSE. Now, $\sum_{j \in \mathcal{R}} \hat{t}^*(j)^2$ is usually close to $\sum_{j \in \mathcal{R}} \hat{t}(j)^2$, since $\hat{t}(j)$ and $\hat{t}^*(j)$ differ only at a small number of points. Furthermore, the terms $\sum_{j \in \mathcal{R}} t(j)^2$ in SSE and $\sum_{j \in \mathcal{R}} \tilde{t}(j)^2$ in $\widehat{\mathrm{SSE}}$ are independent of the smoothing parameter (see (3.2.16) and (3.2.23) below) and will therefore not affect a minimisation of SSE or $\widehat{\mathrm{SSE}}$ over δ. This leaves the cross-product term

$$\sum_{j \in \mathcal{R}} \hat{t}^*(j)\tilde{t}(j)$$

as the one for which we have to find a good estimate.

A. The naive cross-validation method. For $j \in \mathcal{R}_n$, we divide the data $\{X_k : k \in \mathcal{R}_n\}$ into the two parts $\{X_j\}$ and $\{X_k : k \neq j\}$ and define $\tilde{t}_1(j)$ and $\hat{t}_1^*(j)$ by

$$\begin{aligned} \tilde{t}_1(j) &= X_j = (h \star t)(j) + \epsilon_j \\ \hat{t}_1^*(j) &= (2\pi)^{-d} \int_\Theta \{\xi_1^j(\theta)/\chi(\theta)\} e^{-i\langle j,\theta\rangle}\, d\theta, \end{aligned} \tag{3.2.16}$$

where

$$\xi_1^j(\theta) = \sum_{k \neq j} X_k e^{i\langle k,\theta\rangle} + (2d)^{-1} \sum_{|k|=1} X_{j+k} e^{i\langle j,\theta\rangle}. \tag{3.2.17}$$

In (3.2.16), it is clear that $\tilde{t}_1$ copies the technique used for cross-validation in nonparametric regression, and that $\hat{t}_1^*(j)$ is based on the observations $\{X_k : k \neq j\}$. However, unlike the case of nonparametric regression, we cannot leave out the jth observation X_j in the construction of $\hat{t}_1^*(j)$, but have to replace it by an average of observations $\{X_{j+k} : |k| = 1\}$, since we require that $\sum_{j \in \mathcal{R}} \hat{t}_1^*(j)^2$ be close to $\sum_{j \in \mathcal{R}} \hat{t}(j)^2$. We cannot satisfy this last requirement if the numbers of Fourier components in $\xi(\theta)$ and $\xi_1^j(\theta)$ differ, since $\hat{t}(j)$ and $\hat{t}_1^*(j)$ are obtained by partial Fourier inversion.

From (3.2.16) it follows that $\hat{t}_1^*(j)$ and $\hat{t}(j)$ are related in the following way:

$$\begin{aligned}
\hat{t}_1^*(j) &= (2\pi)^{-d}\int_\Theta\{\sum_k X_k e^{i\langle k,\theta\rangle} + (2d)^{-1}\sum_{|k|=1}(X_{j+k}-X_j)e^{i\langle j,\theta\rangle}\}\chi(\theta)^{-1}e^{-i\langle j,\theta\rangle}\,d\theta \\
&= (2\pi)^{-d}\int_\Theta\{\xi(\theta)/\chi(\theta)\}e^{-i\langle j,\theta\rangle}\,d\theta \\
&\quad +(2d)^{-1}\sum_{|k|=1}(X_{j+k}-X_j)(2\pi)^{-d}\int_\Theta \chi(\theta)^{-1}\,d\theta \\
&= \hat{t}(j)+(2d)^{-1}k_\Theta(0)\sum_{|k|=1}(X_{j+k}-X_j), \qquad (3.2.18)
\end{aligned}$$

where

$$k_\Theta(j) = (2\pi)^{-d}\int_\Theta \chi(\theta)^{-1}e^{-i\langle j,\theta\rangle}\,d\theta, \qquad (3.2.19)$$

in analogy with (2.4.5).

Since $(2d)^{-1}k_\Theta(0)\sum_{|k|=1}(X_{j+k}-X_j)$ is small in comparison with $\hat{t}(j)$, $\sum_{j\in\mathcal{R}}\hat{t}_1^*(j)^2$ represents a good approximant to $\sum_{j\in\mathcal{R}}\hat{t}(j)^2$.

B. The deconvolved cross-validation method. In this method we make use of the specific form of the point spread function h as given in (3.2.8). Let $h^{(-1)}$ denote the convolution inverse of h, that is,

$$(h\star h^{(-1)})(j) = \begin{cases} 1 & \text{if } j=0 \\ 0 & \text{otherwise.}\end{cases}$$

Then $h^{(-1)}$ is given by

$$h^{(-1)}(j) = (2\pi)^{-d}\int_\Omega \chi(\theta)^{-1}e^{-i\langle j,\theta\rangle}\,d\theta. \qquad (3.2.20)$$

Now, (3.2.8) implies that

$$h^{(-1)}(j) = \prod_{l=1}^{d}\eta(j_l), \qquad (3.2.21)$$

where

$$\eta(l) = \begin{cases} \lambda^2(1-2\lambda^{-1}+\lambda^{-2}) & \text{if } l=0 \\ \lambda^2(\lambda^{-1}-1) & \text{if } |l|=1 \\ 0 & \text{otherwise.}\end{cases} \qquad (3.2.22)$$

From (3.2.21) and (3.2.22) it follows that $h^{(-1)}$ has support in $\mathcal{N}$, given in (3.2.24) below. Furthermore, $h^{(-1)} = k_\Omega$, as a comparison with (3.2.19) shows.

For $j\in\mathcal{R}_n$, we partition the data $\{X_k : k\in\mathcal{R}_n\}$ in the following way:

$$\{X_{j+\ell} : \ell\in\mathcal{N}\} \quad\text{and}\quad \{X_{j+\ell} : \ell\in\mathbb{Z}^d\setminus\mathcal{N}\},$$

and define $\tilde{t}_2(j)$ and $\hat{t}_2^*(j)$ by

$$\begin{aligned}
\tilde{t}_2(j) &= (X \star h^{(-1)})(j) = t(j) + (\epsilon \star h^{(-1)})(j), \\
\hat{t}_2^*(j) &= (2\pi)^{-d} \int_\Theta \{\xi_2^j(\theta)/\chi(\theta)\} e^{-i\langle j,\theta\rangle}\, d\theta, \qquad (3.2.23)
\end{aligned}$$

where

$$\begin{aligned}
\xi_2^j(\theta) &= \sum_{k \notin j+\mathcal{N}} X_k e^{i\langle k,\theta\rangle} + \sum_{k\in\mathcal{N}} c_{\mathcal{M}}^{-1} \sum_{l\in\mathcal{M}} X_{j+l} e^{i\langle j+k,\theta\rangle}, \\
\mathcal{N} &= \{k \in \mathbb{Z}^d \ : \ \|k\|_\infty \le 1\}, \\
\mathcal{M} &= \{k \in \mathbb{Z}^d \ : \ \|k\|_\infty = 2\}, \quad \text{and} \\
c_{\mathcal{M}} &= |\mathcal{M}|. \qquad (3.2.24)
\end{aligned}$$

Clearly $\tilde{t}_2(j)$ and $\hat{t}_2^*(j)$ are uncorrelated, since $\mathcal{N}$ and $\mathcal{M}$ have no points in common. Furthermore, $\hat{t}_2^*(j)$ and $\hat{t}(j)$ are related by

$$\begin{aligned}
\hat{t}_2^*(j) &= (2\pi)^{-d} \int_\Theta \{\sum_k X_k e^{i\langle k,\theta\rangle} + \sum_{k\in\mathcal{N}} (c_{\mathcal{M}}^{-1} \sum_{l\in\mathcal{M}} X_{j+l} - X_{j+k}) e^{i\langle j+k,\theta\rangle}\} \chi(\theta)^{-1} e^{-i\langle j,\theta\rangle}\, d\theta \\
&= \hat{t}(j) + \sum_{k\in\mathcal{N}} (c_{\mathcal{M}}^{-1} \sum_{l\in\mathcal{M}} X_{j+l} - X_{j+k})(2\pi)^{-d} \int_\Theta \chi(\theta)^{-1} e^{i\langle k,\theta\rangle}\, d\theta \\
&= \hat{t}(j) + \sum_{l\in\mathcal{N}} k_\Theta(l)\ \mathsf{D}_l X_j \qquad (3.2.25)
\end{aligned}$$

with

$$\mathsf{D}_l X_j = c_{\mathcal{M}}^{-1} \sum_{k\in\mathcal{M}} X_{j+k} - X_{j+l},$$

and k_Θ as in (3.2.19).

As in the case of $\hat{t}_1^*(j)$, it is reasonable to assume that the second term in (3.2.25) is small compared to $\hat{t}(j)$, making $\sum_{j\in\mathcal{R}} \hat{t}_2^*(j)^2$ therefore a good approximation to $\sum_{j\in\mathcal{R}} \hat{t}(j)^2$.

In both methods of cross-validation suggested above, $\tilde{t}_i$ and $\hat{t}_i^*$ $(i = 1, 2)$ are defined from complementary subsets of the data. This is crucial to the performance of cross-validation as it implies that $\tilde{t}_i$ and $\hat{t}_i^*$ are independent (just as t and $\hat{t}$ are).

3.3 Results

In the remainder of this chapter we regard δ, the smoothing parameter, and λ, the blur parameter, as functions of n which satisfy

$$\delta \to 0, \quad n\delta \to \infty, \quad \lambda \to \infty, \quad \text{as } n \to \infty. \qquad (3.3.1)$$

In general we shall also assume that $\lambda > 1$. This corresponds to excluding the blur-free case.

We call δ the *optimal smoothing parameter (in a mean sum of squared error sense)* if δ minimises MSSE (see also (1.4.9)). We shall denote the optimal smoothing parameter by δ_m. If $\delta = \arg\min \mathrm{MSSE}$, then the estimator $\hat{t}$ defined on Θ is called the *optimal estimator* for t. For given n and λ let $\delta_0 = \delta_0(n, \lambda)$ denote a smoothing parameter which *minimises the order* of MSSE. For our purpose it will often suffice to work with δ_0 instead of the optimal smoothing parameter δ_m, since $\delta_m = c_m\delta_0$ for some $c_m > 0$, and since we are primarily interested in rates in our asymptotic results.

Before stating our results, we list the assumptions on t, h and ϵ. This list represents a short summary of Subsection 3.2.1. Fix $n > 0$ and write $\mathcal{R}$ for $\mathcal{R}_n$.

A1 The image t is of the form $t = \mathrm{IFT}(\tau)$ with

$$\tau(\theta) \asymp n^d(1 + \|n\theta\|)^{-a} \quad \text{for } \theta \in \Omega = [-\pi, \pi]^d, \quad a > 0.$$

A2 The point spread function h and its Fourier transform χ are given by

$$\begin{aligned} h(j) &= (2\lambda - 1)^{-d}(1 - \lambda^{-1})^{|j|} \quad \text{for } j \in \mathbb{Z}^d, \quad \lambda \geq 1, \\ \chi(\theta) &= \prod_{i=1}^{d}\{1 + 2\lambda^2(1 - \lambda^{-1})(1 - \cos\theta_i)\}^{-1} \quad \text{for } \theta \in \Omega. \end{aligned}$$

A3 The data X are of the form

$$X_j = (h \star t)(j) + \epsilon_j \quad \text{for } j \in \mathbb{Z}^d,$$

where the ϵ_j are independent and identically distributed with mean zero and finite variance σ^2.

3.3.1 Properties of MSSE and SSE

For given image t, blur h, noise ϵ and smoothing set Θ, the partial Fourier inversion estimator $\hat{t}$ of (3.2.12) can be re-written as

$$\begin{aligned} \hat{t}(j) &= (2\pi)^{-d}\int_\Theta \{\tau(\theta) + \nu(\theta)/\chi(\theta)\}e^{-i\langle j,\theta\rangle}\, d\theta \\ &= t_\Theta(j) + (k_\Theta \star \epsilon)(j) \quad \text{for } j \in \mathbb{Z}^d, \end{aligned} \tag{3.3.2}$$

where

$$t_\Theta(j) = (2\pi)^{-d}\int_\Theta \tau(\theta)e^{-i\langle j,\theta\rangle}\, d\theta,$$

and ν denotes the Fourier transform of ϵ.

Assume that t, h and ϵ satisfy **A1–A3**. For $j \in \mathbb{Z}^d$ the bias $\mathcal{B}(j)$ and the variance $\mathcal{V}(j)$ of $\hat{t}(j)$ are

$$\begin{aligned} \mathcal{B}(j) &= -(2\pi)^{-d} \int_{\Omega\setminus\Theta} \tau(\theta) e^{-i\langle j,\theta\rangle}\, d\theta, \\ \mathcal{V}(j) &= \mathbf{E}\{(k_\Theta \star \epsilon)(j)^2\}. \end{aligned} \tag{3.3.3}$$

Assuming that $K = 1$ in (3.2.13), the variance V on $\mathcal{R}$ becomes

$$\begin{aligned} V &\equiv \sum_{j\in\mathcal{R}} \mathcal{V}(j) \\ &= \mathbf{E}\{\sum_{j\in\mathcal{R}} (k_\Theta \star \epsilon)(j)^2\} \\ &= \sum_{j\in\mathcal{R}} \sum_{l} \sum_{m} \mathbf{E}(\epsilon_l \epsilon_m) k_\Theta(j-l) k_\Theta(j-m) \\ &= \sigma^2 \sum_{j\in\mathcal{R}} \sum_{l} k_\Theta(j-l)^2 \\ &= n^d \sigma^2 \sum_{l} k_\Theta(l)^2 \\ &= \sigma^2 (2\pi)^{-d} n^d \int_\Theta \chi(\theta)^{-2}\, d\theta, \end{aligned} \tag{3.3.4}$$

where the last equality follows from Parseval's identity (1.3.9).

The mean sum of squared error, MSSE, on $\mathcal{R}$ is now given by

$$\begin{aligned} \text{MSSE} &= \sum_{j\in\mathcal{R}} \{\mathcal{B}(j)^2 + \mathcal{V}(j)\} \\ &= -(2\pi)^{-2d} \sum_{j\in\mathcal{R}} \{\int_{\Omega\setminus\Theta} \tau(\theta) e^{-i\langle j,\theta\rangle}\, d\theta\}^2 + \sigma^2 (2\pi)^{-d} n^d \int_\Theta \chi(\theta)^{-2}\, d\theta, \end{aligned} \tag{3.3.5}$$

and we now obtain the following result.

Proposition 3.1 *Assume that t, h and ϵ satisfy* **A1–A3** *and that $a > d/2$. If $\lambda\delta \to c$ where $0 \le c \le \infty$ as $n \to \infty$, then, as $n \to \infty$,*

$$MSSE \asymp n^d[(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}].$$

The proof of Proposition 3.1 is given in Subsection 3.4.2.

For the two separate cases $\lambda\delta \to c < \infty$ and $\lambda\delta \to \infty$ one can determine δ_0 as a function of n and λ, where δ_0 is the smoothing parameter which minimises the order of MSSE (see the beginning of this section).

If $\lambda\delta \to c < \infty$, then the conclusion of Proposition 3.1 reduces to

$$\text{MSSE} \asymp n^d\{(n\delta)^{d-2a} + \delta^d\}, \tag{3.3.6}$$

and thus $\delta_0 = n^{-1+d/2a}$ minimises the order of MSSE. On the other hand, if $\lambda\delta \to \infty$ as $n \to \infty$, then MSSE becomes

$$\text{MSSE} \asymp n^d\{(n\delta)^{d-2a} + (\delta^5\lambda^4)^d\}. \tag{3.3.7}$$

The order of MSSE in (3.3.7) is minimised by $\delta_0 = (n^{d-2a}\lambda^{-4d})^{1/(4d+2a)}$, since δ_0 satisfies $(n\delta)^{d-2a} = (\delta^5\lambda^4)^d$. Let MSSE($\delta$) denote the value of MSSE at δ, then

Corollary 3.2 *Assume the conditions of Proposition 3.1.*

1. *If $\lambda\delta \to c < \infty$, and*
$$\delta_0 = n^{-1+d/2a},$$
then δ_0 minimises the order of MSSE amongst δ such that $\lambda\delta \to c$, and
$$MSSE(\delta_0) \asymp n^{d^2/2a}, \quad \text{as } n \to \infty.$$

2. *If $\lambda\delta \to \infty$, then*
$$\delta_0 = \left(n^{d-2a}\lambda^{-4d}\right)^{1/(4d+2a)}$$
minimises the order of MSSE, and
$$MSSE(\delta_0) \asymp n^d(n^{-5}\lambda^4)^{d(2a-d)/(4d+2a)}, \quad \text{as } n \to \infty.$$

Proposition 3.1 shows how MSSE depends on the variables n, δ and λ and on the fixed numbers a and d. The term $n^d(n\delta)^{d-2a}$ derives from squared bias, and variance of $\hat{t}$ produces $(n\delta)^d \max\{1, (\lambda\delta)^{4d}\}$. As one can see from this result, bias decreases as the true image becomes smoother. As far as the variance term is concerned, one has to distinguish two cases: $\lambda\delta \to c < \infty$ and $\lambda\delta \to \infty$ as $n \to \infty$. In Theorems 3.5 and 3.6 below, we shall see that the first case is the only one for which cross-validation works. In this case, MSSE is given by the simple formula (3.3.6). The requirement $\lambda\delta \to c$ puts bounds on λ as can be seen in

Corollary 3.3 *Assume the conditions of Proposition 3.1.*

1. *Assume that $\lambda\delta \to c < \infty$. If δ_0 minimises the order of MSSE over δ such that $\lambda\delta \to c < \infty$, then, as $n \to \infty$,*
$$\lambda = O(n^{1-d/2a}).$$

2. *Assume that $\lambda\delta \to \infty$. If δ_0 minimises the order of MSSE over δ such that $\lambda\delta \to \infty$, then, as $n \to \infty$,*
$$n^{1-d/2a} = o(\lambda) \quad \text{and } \lambda = o(n^{5/4}).$$

It is of interest to compare the rates of MSSE obtained above with those for nonparametric regression. In the case $d = 1$, $a = 2$ one gets the following. If δ_m minimises MSSE and $\lambda\delta_m \to c < \infty$, then $\text{MSSE}(\delta_m) \sim c_1 n^{1/4}$, while the optimal bandwidth h_0 in nonparametric regression leads to $\text{MSSE}(h_0) \sim c_2 n^{1/5}$ for constants c_1, $c_2 > 0$ (see e.g. p131, Theorem 4.1 of Eubank (1988)).

Next we compare SSE and MSSE and show that, to first order, the relative difference between SSE and MSSE becomes negligible.

Proposition 3.4 *Assume that t, h and ϵ satisfy* **A1–A3**. *Assume further that $a > d/2$ and $\mathbf{E}\epsilon^4 < \infty$. If $\lambda\delta \to c$ where $0 \le c \le \infty$ as $n \to \infty$, then, as $n \to \infty$,*

$$\mathbf{E}(SSE - MSSE)^2 = o(MSSE^2).$$

For a proof of this proposition see Subsection 3.4.3.

3.3.2 Mean Square Results for the Cross-Validation Methods

We now turn to the performance of the two cross-validation methods described in Subsection 3.2.3. We shall compare $\widehat{\text{SSE}}_i$ $(i = 1, 2)$ with SSE and show that the $\widehat{\text{SSE}}_i$ are good approximations to SSE in the sense that

$$\widehat{\text{SSE}}_i = \text{SSE} + N_i + o(\text{MSSE}) \qquad \text{in mean square,}$$

where N_i denotes those terms in $\widehat{\text{SSE}}_i - \text{SSE}$ which are independent of the smoothing parameter. The terms in N_i may be large, but as they do not depend on δ, they can be ignored as far as the minimisation of $\widehat{\text{SSE}}_i$ with respect to δ is concerned. Only in this sense, that is, by omitting N_i, can we show that our cross-validation methods perform asymptotically optimally. This situation also occurs in standard cross-validation procedures (see Stone (1984)).

In (3.2.15) $\widehat{\text{SSE}}_i$ was given by

$$\begin{aligned} \widehat{\text{SSE}}_i &= \sum_{j\in\mathcal{R}} \{\hat{t}_i^*(j) - \tilde{t}_i(j)\}^2 \\ &= \sum_{j\in\mathcal{R}} \{\hat{t}_i^*(j)^2 - 2\hat{t}_i^*(j)\tilde{t}_i(j) + \tilde{t}_i(j)^2\} \quad (i = 1, 2). \end{aligned}$$

As indicated in Subsection 3.2.3, the cross-validation estimates $\hat{t}_i^*(j)$ are good approximations to $\hat{t}(j)$, and $\tilde{t}_i(j)$ and $t(j)$ are independent of δ. Instead of considering $\widehat{\text{SSE}}_i$, we may therefore consider

$$\mathrm{SSE}_i^* = \sum_{j\in\mathcal{R}} \{\hat{t}(j)^2 - 2\hat{t}_i^*(j)\tilde{t}_i(j) + t(j)^2\} \quad (i = 1, 2) \tag{3.3.8}$$

as our cross-validation approximation to SSE. Put

$$\begin{aligned} N_1 &= 2\sum_{j\in\mathcal{R}} [t(j)\{\tilde{t}_1(j) - t(j)\} + \{(h \star t)(j) - t(j)\}(\epsilon \star h^{(-1)})(j)], \\ N_2 &= 2\sum_{j\in\mathcal{R}} t(j)(\epsilon \star h^{(-1)})(j), \end{aligned} \tag{3.3.9}$$

and define CV_i by

$$\mathrm{CV}_i = \mathrm{SSE}_i^* + N_i \quad (i = 1, 2). \tag{3.3.10}$$

Clearly, the N_i are independent of δ. In view of the above comments concerning terms that do not depend on δ, showing that

$$\mathbf{E}(\mathrm{CV}_i - \mathrm{SSE})^2 = o(\mathrm{MSSE}^2)$$

therefore is tantamount to showing that the two methods of cross-validation work.

We first turn to the 'naive' cross-validation method which was described in paragraph A of Subsection 3.2.3.

Theorem 3.5 *Assume that t, h and ϵ satisfy* **A1–A3**. *Assume further that $a > 5d/2$, $\lambda > 1$ and $\mathbf{E}\epsilon^4 < \infty$. If δ_m minimises MSSE over δ for given n and λ, then 1 and 2 below are equivalent:*

1. $\mathbf{E}\{(CV_1 - SSE)(\delta_m)^2\} = o\{MSSE(\delta_m)^2\}$;
2. $\lambda = o(n^{1-d/2a})$.

A proof of this theorem is given in Subsection 3.4.4.

The cross-validation method treated in Theorem 3.5 ignores the presence of blur in the image. If one takes careful account of the blur in the image before calculating the cross-validation estimate, one obtains the following result.

Theorem 3.6 *Assume that t, h and ϵ satisfy* **A1–A3**. *Assume further that $a > 5d/2$, $\lambda > 1$ and $\mathbf{E}\epsilon^4 < \infty$. Then 1 and 2 below are equivalent:*

1. $\mathbf{E}(CV_2 - SSE)^2 = o(MSSE^2)$;
2. $(n^{-1}\lambda^4\delta^2)^d\{1 + (\lambda\delta)^{4d}\}[(n\delta)^{d-2a} + \delta^d\{1 + (\lambda\delta)^{4d}\}]^{-2} \to 0$.

If δ_m minimises MSSE over δ for given n and λ, and if $\delta = c_3\delta_m$ for $c_3 > 0$, then 2 holds if and only if $\lambda = o(n^{1/4})$.

A proof of Theorem 3.6 can be found in Subsection 3.4.5.

We conclude this subsection with some comments about our results.

3.3.3 Discussion of Results

Asymptotic performance of cross-validation with respect to MSSE. In Theorems 3.5 and 3.6 we compared the performance of CV_i with SSE. In connection with Proposition 3.4, these results now also imply that

$$\text{CV}_i = \text{MSSE}\{1 + o(1)\} \quad \text{as } n \to \infty \tag{3.3.11}$$

in mean square, and hence in probability.

Ceilings for the permissible amount of blur. The theorems show that both methods of cross-validation perform asymptotically optimally provided the blur is not too large. For the second method the blur ceiling is described by the condition $\lambda = o(n^{1/4})$, while $\lambda = o(n^{1-d/2a})$ describes the ceiling for the first method. The latter blur ceiling compares favourably with the blur ceiling for MSSE in the case where $\lambda\delta \to c < \infty$ (see Corollary 3.3).

Implications of the blur ceiling for the model. The blur ceiling for both cross-validation methods is rather low: it grows at a much slower rate than n if cross-validation is to provide asymptotic minimisation of SSE or MSSE. Thus we are considering models in which optimal image enhancement is achievable by cross-validation only if the amount of blur in a scene decreases as the scene is recorded in more detail (see also paragraph C of Subsection 3.2.1).

Comparison of the blur ceiling for the two cross-validation methods. For $a > 5d/2$ one can compare the two blur ceilings. Since then $1 - d/2a > 1/4$, the blur ceiling of the first method is higher than that of the second method. This means that the first method allows a wider range of blur than the second method, and therefore offers a greater range of application. This may seem rather unexpected in view of the fact that careful account has been taken of the blur in the second method. However, by deconvolving with the blur, as is done in the second method, the noise component — and therefore also the variance — increases dramatically. To understand why this happens, observe that $\hat{t}$ is calculated on a very small smoothing set Θ as $n \to \infty$. The approximation $\tilde{t}_2$ to t, however is of the form

$$\tilde{t}_2 = \hat{t}_\Omega = X \star h^{(-1)},$$

and the non-zero components of $h^{(-1)}$ grow like λ^{2d} (see (3.2.22)). The random part of $\tilde{t}_2$, namely $\epsilon \star h^{(-1)}$, therefore also grows like λ^{2d}, and $\lambda \to \infty$ as $n \to \infty$.

Comparison of CV_1 and SSE for large amounts of blur. The limitations $\lambda = o(n^{1-d/2a})$ and $\lambda = o(n^{1/4})$ are genuine restrictions which must be enforced to ensure that optimal performance is achievable. In fact, the proof of Theorem 3.5 can be extended to show that

$$\begin{aligned} \mathbf{E}(\mathrm{CV}_1 - \mathrm{SSE})^2/\mathrm{MSSE}^2 &\asymp 1 \quad \text{if } \lambda \sim const.n^{1-d/2a}; \text{ and} \\ \mathbf{E}(\mathrm{CV}_1 - \mathrm{SSE})^2/\mathrm{MSSE}^2 &\to \infty \quad \text{if } n^{1-d/2a} = o(\lambda). \end{aligned}$$

Corresponding results also apply to the second method.

Dependence of CV_2 on the point spread function. The second cross-validation method makes explicit use of the finite support $\mathcal{N}$ of $h^{(-1)}$. For example, for each $j \in \mathcal{R}_n$, the data is partitioned into $\{X_{j+k} \ : \ k \notin \mathcal{N}\}$ and $\{X_{j+k} \ : \ k \in \mathcal{N}\}$. The former set is used to construct $\hat{t}_2^*$, and the latter to construct $\tilde{t}_2$. This method is therefore more restrictive in its applicability to other classes of point spread functions than the first method, which has an obvious extension to other classes of point spread functions.

Comparison of the two cross-validation methods. The comments and comparisons of the two methods seem to indicate that for the relatively low levels of blur for which both cross-validation methods work, the naive method is preferable in more than one way. This is confirmed by numerical results presented in Subsection 3.3.5.

3.3.4 Uniform *a.s.* Results for the Naive CV Method

Proposition 3.4, Theorems 3.5 and 3.6 show that SSE, CV_1 and CV_2 converge to MSSE in a mean square sense. They imply that

$$(\mathrm{SSE} - \mathrm{MSSE})/\mathrm{MSSE} \to 0 \qquad \textit{in pr.}; \tag{3.3.12}$$

$$(\mathrm{CV}_i - \mathrm{SSE})(\delta)/\mathrm{MSSE}(\delta) \to 0 \qquad \textit{in pr.}, \quad \text{for } i = 1, 2, \tag{3.3.13}$$

where $\delta = c_3\delta_m$, δ_m denotes the minimiser of MSSE and $c_3 > 0$. These properties can be extended in two ways: one can show that (3.3.13) holds for a range of values of the smoothing parameter δ, and that convergence in (3.3.12) and (3.3.13) can be strengthened to *a.s.* convergence.

In this subsection we exhibit sufficient conditions for t, λ and ϵ which allow us to show that

$$\sup_{\delta \in I} |(\mathrm{CV}_1 - \mathrm{MSSE})(\delta)|/\mathrm{MSSE}(\delta) \to 0 \qquad a.s., \tag{3.3.14}$$

where I denotes an interval which contains the minimiser δ_m of MSSE.

Similar results may be obtained for the second cross-validation method, but no new insights would be gained by doing this, as the same techniques are used to prove either result.

The proof of (3.3.14) is accomplished in three steps: on the interval I we define a grid; the grid points are equally spaced and separated by $n^{-\gamma}$ for some $\gamma > 0$. We first show that for any two points ρ and ϕ in I, ρ a grid-point and $|\rho - \phi| \leq n^{-\gamma}$, the absolute value of the differences $|\mathrm{CV}_1 - \mathrm{SSE}|/\mathrm{MSSE}$ is at most $c_1 n^{-1}$ for some $c_1 > 0$. In the second step, accomplished in Theorem 3.8, we show that

$$\sup_{\rho_i \in I} |(\mathrm{CV}_1 - \mathrm{SSE})(\rho_i)|/\mathrm{MSSE}(\rho_i) \to 0 \qquad a.s.$$

To obtain a generalisation of (3.3.11) to *a.s.* convergence, one also has to show that $\mathrm{SSE} - \mathrm{MSSE} = o(\mathrm{MSSE})$ *a.s.* uniformly on I. This is the purpose of Proposition 3.9. Sufficient conditions for our final conclusions are presented in Theorem 3.10, which is a straightforward consequence of Propositions 3.7, 3.9 and Theorem 3.8.

In some of the results of this subsection, we shall make use of the following assumptions on t and h.

A4 The blur parameter λ grows with n like $\lambda = O(n^{1-\eta-d/2a})$ for some $\eta > 0$.

A5 For $0 < \Delta \leq d^{-1}$, $|t(j+k) - t(j)| = O(n^{-\Delta d})$ for $j, k \in \mathcal{R}_n$, $|k| \leq 1$.

Assumption **A4** implies that $\lambda = o(n^{1-d/2a})$, the condition required in Theorem 3.5; and **A5** expresses a Lipschitz condition of order Δd for t. In the case of differentiable functions with bounded first derivative, a Lipschitz condition of order 1 is always satisfied. Since we assume that $a \geq 1$, **A5** is natural in view of the fact that one may regard t as a discretisation of an a times continuously differentiable function T (see also paragraph A of Subsection 3.2.3).

Let δ_m denote the minimiser of MSSE over δ for given n and λ. Then $\delta_m = c_m \delta_0$, with $\delta_0 = n^{-1+d/2a}$, $c_m > 0$. For $n > 0$, $0 < k_1 < k_2 < \infty$ such that $k_1\delta_0 < \delta_m < k_2\delta_0$, put

$$I = [k_1\delta_0,\ k_2\delta_0]. \tag{3.3.15}$$

On I define grid points ρ_i and points $\phi_{i,r}$ by

$$\begin{aligned} \rho_i &= k_1\delta_0 + in^{-\gamma} \quad \text{for } 0 \le i < \kappa \\ \phi_{i,r} &= \rho_i + rn^{-\gamma} \quad \text{for } 0 \le i < \kappa \text{ and } 0 < r \le 1, \end{aligned} \tag{3.3.16}$$

where $\gamma > 0$ and $\kappa \le k_3 n^{\gamma - 1 + d/2a}$, $k_3 > k_2 - k_1$.

The next proposition shows how to choose γ such that the the quotients $|\mathrm{CV}_1 - \mathrm{SSE}|/\mathrm{MSSE}$, calculated at ρ_i and $\phi_{i,r}$, are at most $c_1 n^{-1}$ for $c_1 > 0$. Recall that assumptions **A1–A3** were listed at the beginning of this section.

Proposition 3.7 *Assume that t, h and ϵ satisfy* **A1–A4**. *Assume further that $a > 3d$ and $\mathbf{E}\epsilon^4 < \infty$. If $\gamma > 2 + 2a - d/2a$ then*

$$\sup_{0 \le i < \kappa} \sup_{0 < r \le 1} \left| \frac{CV_1(n,\rho_i) - SSE(n,\rho_i)}{MSSE(n,\rho_i)} - \frac{CV_1(n,\phi_{i,r}) - SSE(n,\phi_{i,r})}{MSSE(n,\phi_{i,r})} \right| = o(n^{-1}) \quad a.s.$$

A proof of this proposition is given in Subsection 3.4.6.

We are now ready to present the generalisation of Theorem 3.5 to *a.s.* convergence.

Theorem 3.8 *Assume that t, h and ϵ satisfy* **A1–A5** *with $\Delta \ge d/4a$. Assume further that $a > 3d$ and $\gamma > 2 + 2a - d/2a$. If*

1. *$b \ge (\gamma + \omega + d/2a)\max\{2a/d^2, (4\eta)^{-1}\}$ for $\omega > 0$, and*
2. *$\mathbf{E}|\epsilon|^r < \infty$ for $1 \le r \le 2b$,*

then

$$\sup_{\delta \in I} \left| \frac{CV(n,\delta) - SSE(n,\delta)}{MSSE(n,\delta)} \right| \to 0 \quad a.s. \quad as\ n \to \infty.$$

A proof of Theorem 3.8 can be found in Subsection 3.4.7.

A comparison of Theorems 3.5 and 3.8 shows that the latter requires slightly stronger assumptions on image, point spread function and the error distribution. As pointed out above, assumptions **A4** and **A5**, which relate to the point spread function and the image, are not much more restrictive than the assumptions imposed on h and t in Theorem 3.5. As we shall see in the proof of Theorem 3.8, assumption 2 of Theorem 3.8 was made in order to prove the *a.s.* convergence by means of the Borel-Cantelli lemma.

We now turn to the strengthening of Proposition 3.4. Since the methods of proof of our Proposition 3.9 are virtually the same as those employed in the proofs of Proposition 3.7 and Theorem 3.8, we shall only give an outline of a proof containing the essential features. This can be found in Subsection 3.4.8.

Proposition 3.9 *Assume that t, h and ϵ satisfy* **A1–A3** *and that $\lambda = O(n^{1-d/2a})$. Assume further that $a > 3d$ and $\gamma > 2 + 2a - d/2a$. If*

1. $b \geq (\gamma + \omega + d/2a)2a/d$ *for* $\omega > 0$, *and*
2. $\mathbf{E}|\epsilon|^r < \infty$ *for* $1 \leq r \leq 2b$,

then

$$\sup_{\delta \in I} \left| \frac{SSE(n,\delta) - MSSE(n,\delta)}{MSSE(n,\delta)} \right| \to 0 \quad a.s. \quad as\ n \to \infty.$$

In Proposition 3.9 we restricted attention to point spread functions h with $\lambda = O(n^{1-d/2a})$, the rate of growth given in Corollary 3.3. Since we are only interested in point spread functions up to this ceiling of λ, the full analogue of Proposition 3.4 is not given here.

It is interesting to observe that Theorem 3.8 and Proposition 3.9 make use of the same grid I where grid points are separated by $n^{-\gamma}$. In contrast to this, it is worth noticing that the order of the bounded moments required in Theorem 3.8 and Proposition 3.9 is slightly different. Furthermore, the Lipschitz condition **A5** of Theorem 3.8 is not required in Proposition 3.9.

We conclude this subsection with Theorem 3.10, which is now a straightforward corollary of the previous results.

Theorem 3.10 *Assume that t, h and ϵ satisfy* **A1–A5** *with $\Delta \geq d/4a$. Assume further that $a > 3d$ and $\gamma > 2 + 2a - d/2a$. If*

1. $b \geq (\gamma + \omega + d/2a)\max\{2a/d,\ (4\eta)^{-1}\}$ *for* $\omega > 0$, *and*
2. $\mathbf{E}|\epsilon|^r < \infty$ *for* $1 \leq r \leq 2b$,

then

$$\sup_{\delta \in I} \left| \frac{CV_1(n,\delta) - MSSE(n,\delta)}{MSSE(n,\delta)} \right| \to 0 \quad a.s. \quad as\ n \to \infty.$$

□

3.3.5 Numerical Examples

On the interval $[0,1]$ we considered images $T^{(1)}$ and $T^{(2)}$ defined by

$$\begin{aligned} T^{(1)} &= \sin(2\pi x)\sin(2\pi(x-0.3))\sin(2\pi(x-0.5)) \\ T^{(2)} &= \begin{cases} (x-\frac{20}{512})(x-0.3)(x-0.575)(x-\frac{491}{512}) + 0.2 & x \in [\frac{20}{512}, \frac{491}{512}] \\ 0.2 & x \leq \frac{20}{512},\ x \geq \frac{491}{512} \end{cases} \end{aligned}$$

These images were discretised to 512 points, giving rise to the true (discrete) images $t^{(1)}$ and $t^{(2)}$. The solid lines of Figures 3.1 and 3.2 show the graphs of $t^{(1)}$ and $t^{(2)}$ respectively. The true images $t^{(1)}$ and $t^{(2)}$ were convolved with the family of point spread functions $h = \{h_\lambda\}$ (see **A2**) for λ varying from 1 to 80. The blurred images $h \star t$ were further degraded by Gaussian additive noise ϵ of mean zero, unit variance and standard deviation 0.2, leading to observations

$$X^{(i)} = h_\lambda \star t^{(i)} + \epsilon^{(i)} \quad (i = 1, 2).$$

For $\lambda = 40$, the observations $X^{(1)}$ and $X^{(2)}$ are displayed in Figures 3.3 and 3.4 respectively Ideally, all figures should be drawn on the same scale and the x-axis should cover the range $[0, 1]$. The software S however did not easily lend itselef to this.

The Fourier transform ξ of X was calculated at 512 points by means of the Fast Fourier Transform. For smoothing sets Θ, the estimates $\hat{t}^{(1)}$ for $t^{(1)}$ and $\hat{t}^{(2)}$ for $t^{(2)}$ were calculated using the inverse Fast Fourier Transform and (3.2.12). Because of the discrete nature of the Fast Fourier Transform, only integer values could be used for the smoothing parameter δ.

The sum of squared error, SSE, was calculated for $\delta = 1, 2, \ldots, 10$ and $\lambda \in [1, 80]$. For the smoother image $t^{(1)}$, $\delta = 4$ minimised SSE over the range of point spread functions considered, while $\delta = 3$ minimised SSE for $t^{(2)}$ over the same range of point spread functions. ⊕ Insert The graphs of the optimal estimates $\hat{t}^{(1)}$ for $t^{(1)}$ and $\hat{t}^{(2)}$ for $t^{(2)}$ (calculated using $\delta = 4$ for the first and $\delta = 3$ for the second estimate) are given by the lines composed of dots and dashes in Figures 3.1 and 3.2 respectively for $\lambda = 40$.

Note that the estimates $\hat{t}^{(1)}$ and $\hat{t}^{(2)}$ are rather smooth. This is to be expected for the following reason. If one identifies the smoothing set Θ with its indicator function $\mathcal{I}_\Theta$, then one obtains

$$\begin{aligned}
\hat{t}(j) &= (2\pi)^{-d} \int_\Omega \xi(\theta)\chi(\theta)^{-1}\mathcal{I}_\Theta(\theta)e^{-i\langle j,\theta\rangle}\, d\theta \\
&= (X \star h^{(-1)} \star w_\Theta)(j),
\end{aligned}$$

where w_Θ denotes the inverse Fourier transform of $\mathcal{I}_\Theta$. Now observe that if $2L$ is the width of the rectangular window $\mathcal{I}_\Theta$, then $w_\Theta(j) = cj^{-1}\sin(jL)$ for $c > 0$. The larger the number of zeroes and sidelobes of w_Θ that lie inside the region of interest (here defined by $j = 1, \ldots, 512$), the rougher (or noisier) $\hat{t}$ becomes. Similarly, $\hat{t}$ becomes smoother with decreasing numbers of zeroes and sidelobes inside $j = 1, \ldots, 512$. The latter situation applies to our smoothing sets Θ, which are very narrow: they contain 8 points for $\hat{t}^{(1)}$ and 6 points for $\hat{t}^{(2)}$ of the possible 512 Fourier components.

To examine the performance of the two types of cross-validation, we calculated the pairs $(\tilde{t}_i,\ \hat{t}_i^*)$, $i = 1, 2$ as described in Subsection 3.2.3 and the cross-validation approximations CV_1 and CV_2 to SSE over a range of δs for each of the two images $t^{(1)}$ and $t^{(2)}$. For the range of point spread functions considered, the optimal smoothing parameters,

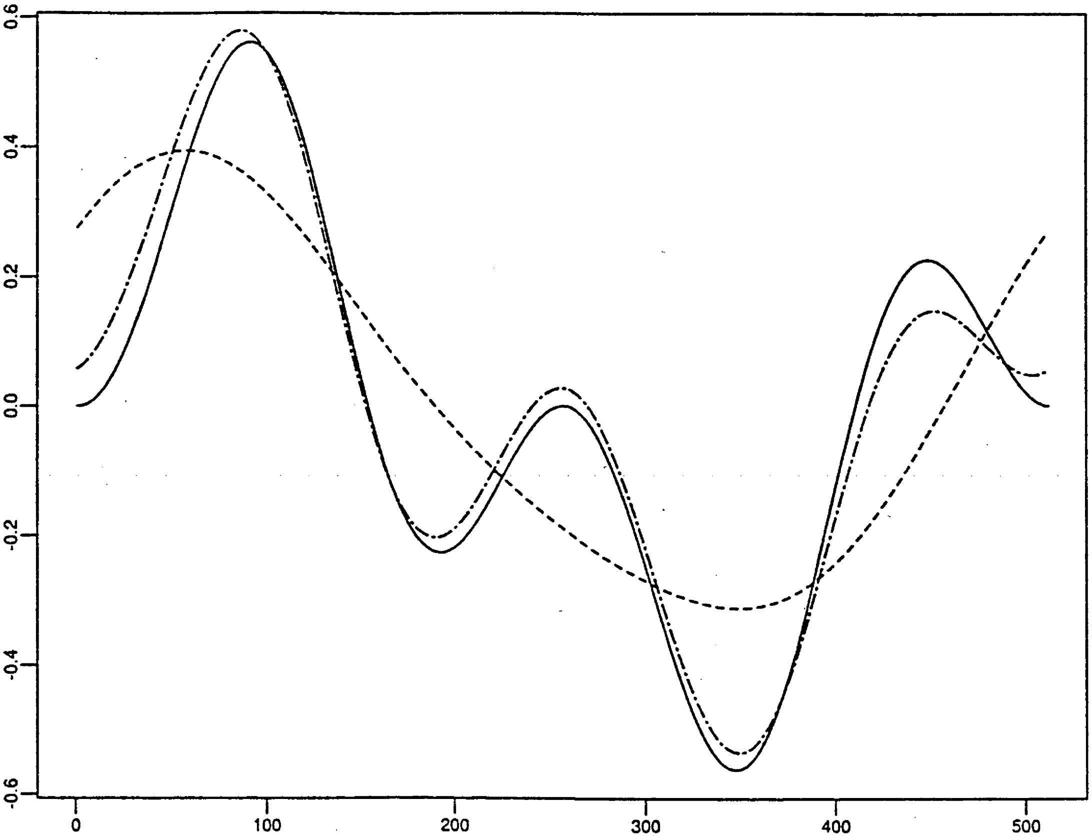

Figure 3.1: True image and estimates for $t^{(1)}$.

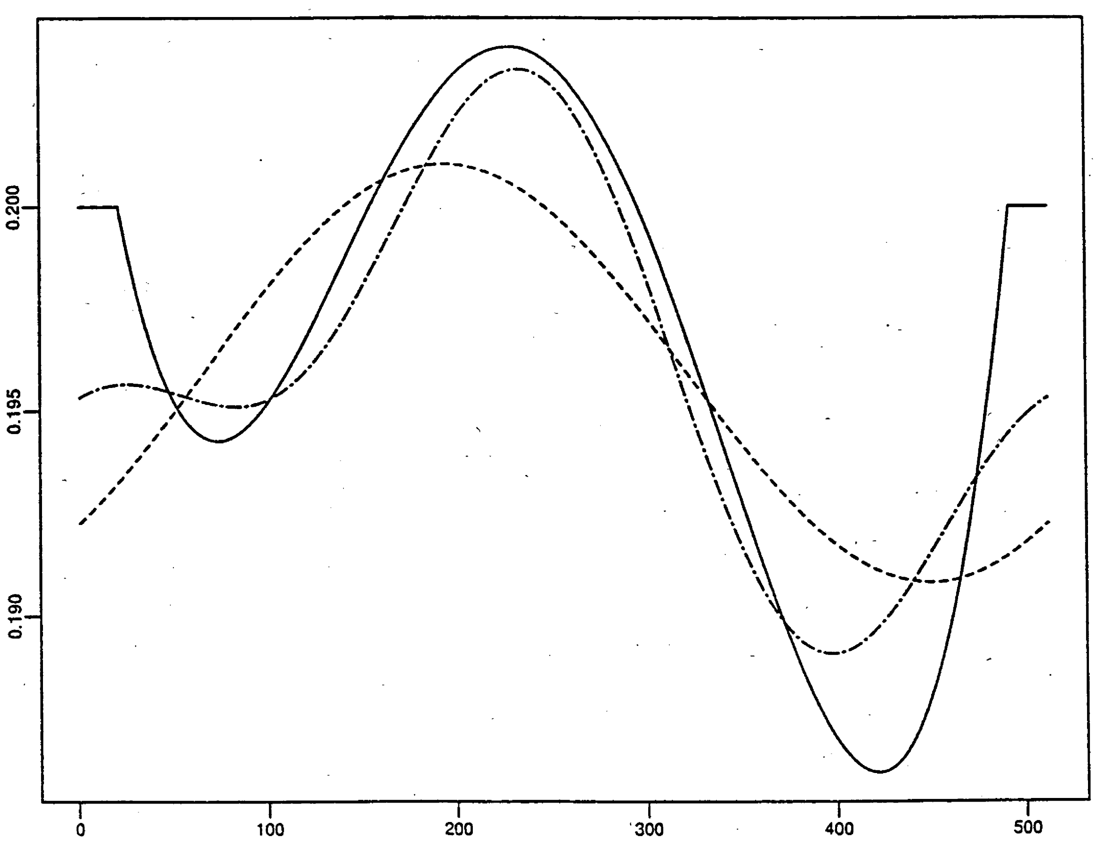

Figure 3.2: True image and estimates for $t^{(2)}$.

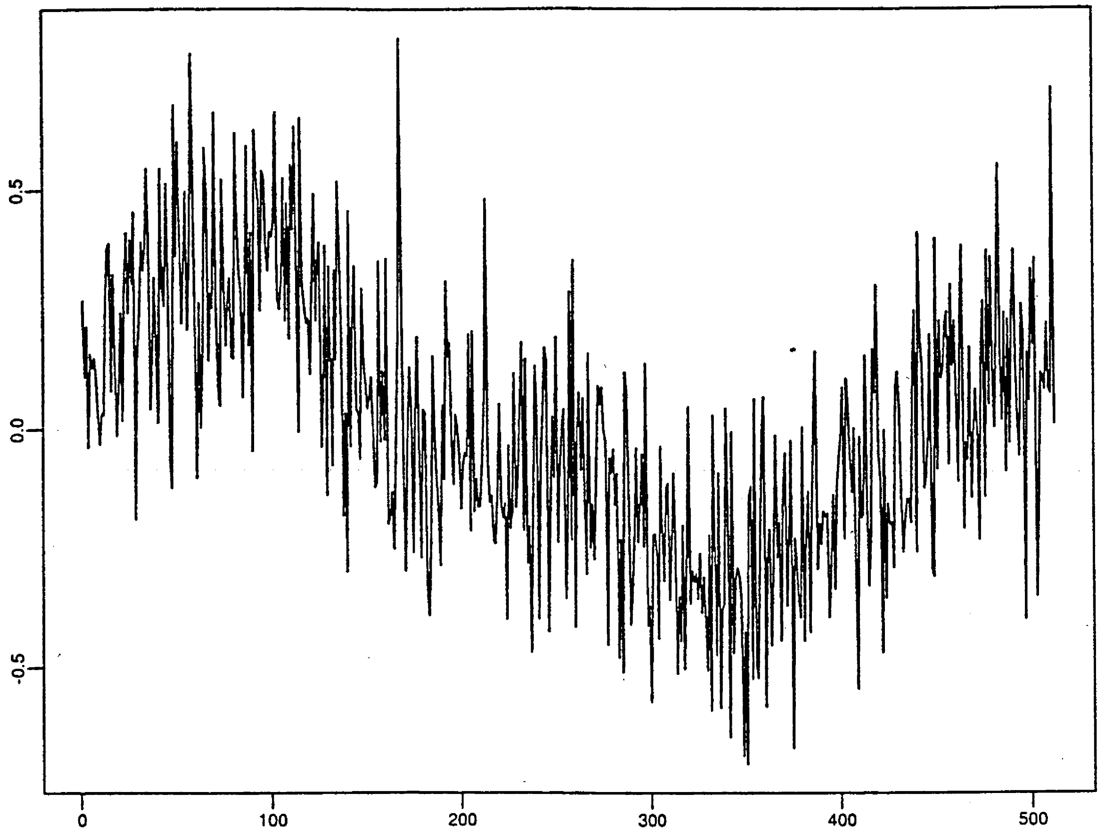

Figure 3.3: Observations of $t^{(1)}$.

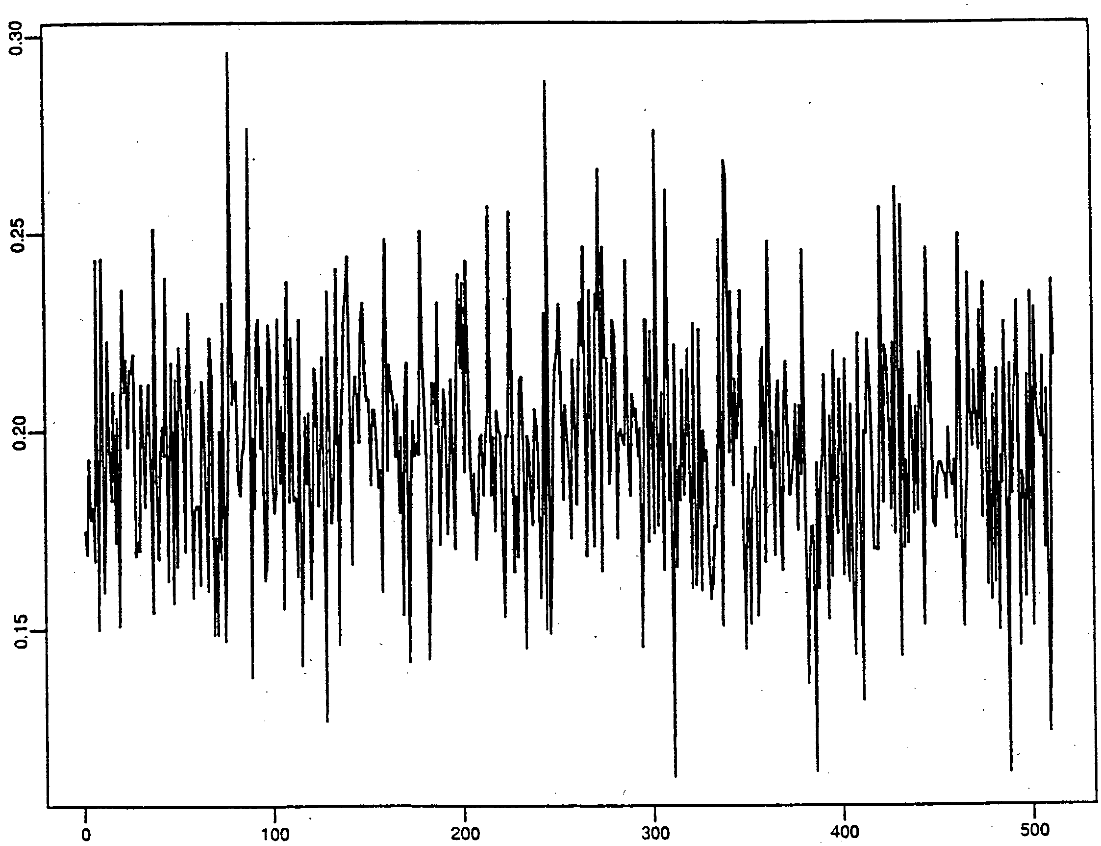

Figure 3.4: Observations of $t^{(2)}$.

namely $\delta = 4$ and $\delta = 3$, minimised CV_1 for $t^{(1)}$ and $t^{(2)}$. This shows that the naive cross-validation method selected the optimal smoothing parameter. The second method did not perform as well: $\delta = 2$ minimised CV_2 for $t^{(1)}$ and $t^{(2)}$. For the smoothing parameters which minimised CV_1 and CV_2 we calculated the estimates $\hat{t}_i$ according to (3.2.12). The graphs of these estimates are displayed in Figures 3.1 and 3.2. For the first cross-validation method, $\hat{t}_1$ agreed with the optimal estimate (since the same value of δ was used). The second cross-validation method selected smaller values of the smoothing parameter. This leads to the oversmoothed estimates $\hat{t}_2$ which are given by the broken lines in Figures 3.1 and 3.2.

The numerical results confirm the theoretical results presented in Subsection 3.3.2. In particular they show that the first cross-validation method performs very well, and is better than the second method for the range of point spread functions and the rough and smooth true images $t^{(2)}$ and $t^{(1)}$ which were considered.

3.4 Proofs

We begin this section by stating and proving some technical lemmas which are used repeatedly in the proofs of the propositions and theorems given in Section 3.3. Lemmas which are used in the proof of one proposition or theorem are stated and proved as part of that proposition or theorem. The lemmas appearing in this section are numbered consecutively as Lemma 3.4.1, Lemma 3.4.2, etc.

3.4.1 Some Technical Lemmas

Lemma 3.4.1 *For $n > 0$, $a > d/k$, $k = 1,2$ and $\tau(\theta) \asymp n^d(1 + \|n\theta\|)^{-a}$, $\theta \in \Omega$, the following hold as $n \to \infty$:*

1. $\int_{\Omega \setminus \Theta} \tau(\theta)^k \, d\theta \asymp n^{(k-1)d}(n\delta)^{d-ka}$;
2. $\int_{\Omega} \tau(\theta)^k \, d\theta \asymp n^{(k-1)d}$.

Proof of Lemma 3.4.1

To prove part 1, put $\mathsf{C}\Theta = \Omega \setminus \Theta$, then

$$\begin{aligned}\int_{\mathsf{C}\Theta} \tau(\theta)^k \, d\theta &\asymp n^{kd} \int_{\mathsf{C}\Theta} (1 + \|n\theta\|)^{-ka} \, d\theta \\ &\asymp n^{(k-1)d} \int_{\mathsf{C}\Theta_n} (1 + \|\theta\|)^{-ka} \, d\theta,\end{aligned}$$

where $C\Theta_n = \{\psi \in \mathbf{R}^d : \psi = n\theta, \theta \in C\Theta\}$. By a further change of variable $\theta \rightsquigarrow (r, \phi_1, \ldots, \phi_{d-1})$ with $r = \|\theta\|$, one now obtains

$$\begin{aligned} \int_{C\Theta} \tau(\theta)^k \, d\theta &\asymp n^{(k-1)d} \int_{n\delta}^{n\pi} (1+r)^{-ka+d-1} \, dr \\ &\asymp n^{(k-1)d} (n\delta)^{d-ka}. \end{aligned}$$

This completes the proof of part 1.

To show part 2, take $\Theta = \{0\}$ in the proof of part 1. It now follows that

$$\begin{aligned} \int_{\Omega} \tau(\theta)^k \, d\theta &\asymp n^{(k-1)d} \int_{0}^{n\pi} (1+r)^{-ka+d-1} \, dr \\ &\asymp n^{(k-1)d}, \end{aligned}$$

since $ka > d$. □

Lemma 3.4.2 *For* $\chi(\theta) = \prod_{i=1}^d \{1 + 2\lambda^2(1-\lambda^{-1})(1-\cos\theta_i)\}^{-1}$, $\lambda > 1$, *one has*

$$\int_{\Theta} \chi(\theta)^{-p} \, d\theta \asymp \delta^d \max\{1, (\lambda\delta)^{2p}\}^d \quad \text{for } p = 1, 2, \ldots \quad \text{as } n \to \infty.$$

Proof of Lemma 3.4.2

If $\omega \in [0, \pi]$, then

$$1 - \cos\omega = 2\sin^2(\omega/2) \asymp \omega^2/2.$$

From this relationship it follows that

$$\chi(\theta) \asymp \prod_{i=1}^d \{1 + \lambda^2(1-\lambda^{-1})\theta_i^2\}^{-1} \asymp \prod_{i=1}^d (1 + \lambda^2\theta_i^2)^{-1},$$

since $\lambda > 1$. Similarly, for $\theta \in \Theta$,

$$\chi(\theta)^{-1} \asymp \prod_{i=1}^d (1 + \lambda^2\theta_i^2)$$

follows, since $\chi(\theta) \neq 0$ on Θ. Hence

$$\begin{aligned} \int_{\Theta} \chi(\theta)^{-p} \, d\theta &\asymp \prod_{i=1}^d \int_0^{\delta} (1 + \lambda^2\theta_i^2)^p \, d\theta_i \asymp \{\int_0^{\delta} (1 + \lambda^2\omega^2)^p \, d\omega\}^d \\ &\asymp \delta^d \{(1 + \lambda^2\delta^2)^p\}^d \asymp \delta^d \max(1, \lambda^{2p}\delta^{2p})^d. \end{aligned}$$

□

Lemma 3.4.3 *For $n > 0$, $a > d/2$ and $j \notin \mathcal{R}$, the following holds as $n \to \infty$:*

$$\sum_{j \notin \mathcal{R}} \mathcal{B}(j)^2 = O[(n\delta)^{2(d-a)} \max_{r=1,\ldots,d} \{n^r (n\delta)^{-2r}\}].$$

Proof of Lemma 3.4.3

Recall that for $n > 0$, $\mathcal{R} = \mathcal{R}_n$ is defined by

$$\mathcal{R} = \{j \in \mathbb{Z}^d \ : \ |j_i| \leq Kn, \ \ i = 1, \ldots, d\}.$$

For $j \in \mathbb{Z}^d \setminus \mathcal{R}$ and $1 \leq r \leq d$, put

$$\mathcal{S}(r) = \{j \in \mathbb{Z}^d \setminus \mathcal{R} \ : \ |j_1|, \ldots, |j_r| > Kn \text{ and } |j_{r+1}|, \ldots, |j_d| \leq Kn\};$$

and put

$$\mathsf{C}\Theta = \Omega \setminus \Theta, \quad \mathsf{C}\Theta_n = \{\psi \in \mathbf{R}^d \ : \ \psi = n\theta, \ \ \theta \in \mathsf{C}\Theta\}.$$

For $j \in \mathcal{S}(r)$ one has

$$\begin{aligned} \mathcal{B}(j) &= (2\pi)^{-d} \int_{\mathsf{C}\Theta} \tau(\theta) \cos\langle j, \theta \rangle \, d\theta \\ &\asymp \int_{\mathsf{C}\Theta_n} (1 + \|\theta\|)^{-a} \cos\langle n^{-1} j, \theta \rangle \, d\theta \\ &\asymp n^r (\prod_{i=1}^r j_i)^{-1} \int_{\mathsf{C}\Theta_n} (1 + \|\theta\|)^{-a} \{\prod_{i=1}^r (\partial / \partial \theta_i)\} (\mathrm{cs}^{(r)} \langle n^{-1} j, \theta \rangle) \, d\theta, \end{aligned} \tag{3.4.1}$$

where $\mathrm{cs}^{(r)} = \pm \cos$ or $\mathrm{cs}^{(r)} = \pm \sin$, depending on r.

We claim that

$$\mathcal{B}(j) = O\{n^r (\prod_{i=1}^r j_i)^{-1} (n\delta)^{d-r-a}\}. \tag{3.4.2}$$

To show (3.4.2) it suffices to estimate the integral in (3.4.1). Note that $\mathsf{C}\Theta_n$ is of the form

$$\mathsf{C}\Theta_n = \bigcup_{p=1}^{l} R_p = \bigcup_{p=1}^{l} \prod_{m=1}^{d} [a_m, b_m]$$

with $a_m, b_m \in \mathbf{R}$. Integration over $\mathsf{C}\Theta_n$ can therefore be decomposed into a sum of integrals over R_p. We use integration by parts on each R_p.

Put

$$\begin{aligned} f(\theta) &= (1 + \|\theta\|)^{-a}; \\ (\partial / \partial \theta_i) g_r(\theta) &= (\partial / \partial \theta_i) \mathrm{cs}^{(r)} \langle n^{-1} j, \theta \rangle. \end{aligned}$$

The first two steps are as follows for $R = R_p$ $(p = 1, \ldots, l)$:

$$\begin{aligned}
\rho_p &\equiv \int_R (1 + \|\theta\|)^{-a} \{\prod_{i=1}^{r} (\partial/\partial\theta_i)\} (\mathrm{cs}^{(r)} \langle n^{-1}j, \theta\rangle)\, d\theta \\
&= \int_R f(\theta) \{\prod_{i=2}^{r} (\partial/\partial\theta_i)\} (\partial/\partial\theta_1) g_r(\theta)\, d\theta \\
&= \int_{a_2}^{b_2} \cdots \int_{a_d}^{b_d} \Big[f(b_1, \theta_2, \ldots, \theta_d) \{\prod_{i=2}^{r} (\partial/\partial\theta_i)\} g_r(b_1, \theta_2, \ldots, \theta_d) \\
&\qquad - f(a_1, \theta_2, \ldots, \theta_d) \{\prod_{i=2}^{r} (\partial/\partial\theta_i)\} g_r(a_1, \theta_2, \ldots, \theta_d) \Big]\, d\theta_2 \ldots d\theta_d \\
&\qquad - \int_R \{(\partial/\partial\theta_1) f(\theta)\} \{\prod_{i=2}^{r} (\partial/\partial\theta_i)\} g_r(\theta)\, d\theta \\
&= \int_{a_3}^{b_3} \cdots \int_{a_d}^{b_d} \Big[f(b_1, b_2, \theta_3, \ldots, \theta_d) \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(b_1, b_2, \theta_3, \ldots, \theta_d) \\
&\qquad - f(b_1, a_2, \theta_3, \ldots, \theta_d) \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(b_1, a_2, \theta_3, \ldots, \theta_d) \\
&\qquad - f(a_1, b_2, \theta_3, \ldots, \theta_d) \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(a_1, b_2, \theta_3, \ldots, \theta_d) \\
&\qquad + f(a_1, a_2, \theta_3, \ldots, \theta_d) \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(a_1, a_2, \theta_3, \ldots, \theta_d) \Big]\, d\theta_3 \ldots d\theta_d \\
&\qquad - \int_{a_2}^{b_2} \cdots \int_{a_d}^{b_d} \Big[\{(\partial/\partial\theta_2) f(b_1, \theta_2, \ldots, \theta_d)\} \\
&\qquad\qquad \times \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(b_1, \theta_2, \ldots, \theta_d) \\
&\qquad\quad - \{(\partial/\partial\theta_2) f(a_1, \theta_2, \ldots, \theta_d)\} \\
&\qquad\qquad \times \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(a_1, \theta_2, \ldots, \theta_d) \Big]\, d\theta_2 \ldots d\theta_d \\
&\qquad - \int_{a_1}^{b_1} \int_{a_3}^{b_3} \cdots \int_{a_d}^{b_d} \Big[\{(\partial/\partial\theta_1) f(\theta_1, b_2, \theta_3, \ldots, \theta_d)\} \\
&\qquad\qquad \times \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(\theta_1, b_2, \theta_3, \ldots, \theta_d) \\
&\qquad\quad - \{(\partial/\partial\theta_1) f(\theta_1, a_2, \theta_3, \ldots, \theta_d)\} \\
&\qquad\qquad \times \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(\theta_1, a_2, \theta_3, \ldots, \theta_d) \Big]\, d\theta_1 d\theta_3 \ldots d\theta_d \\
&\qquad + \int_R \{(\partial^2/\partial\theta_1\partial\theta_2) f(\theta)\} \{\prod_{i=3}^{r} (\partial/\partial\theta_i)\} g_{r-1}(\theta)\, d\theta.
\end{aligned}$$

After r steps one obtains for ρ_p terms of the form

$$\begin{aligned}
\rho_p &= \sum \int \cdots \int (1 + \|\theta\|_{(-r)})^{-a} \sin\langle n^{-1}j, \theta\rangle_{(-r)}\, d\theta_{r+1} \ldots d\theta_d \\
&\qquad + c_1 \sum \int \cdots \int (1 + \|\theta\|_{(-r+1)})^{-a-2} \theta_i \sin\langle n^{-1}j, \theta\rangle_{(-r+1)}\, d\theta_r \ldots d\theta_d
\end{aligned}$$

$$+\ldots+c_r\sum\int_R(1+\|\theta\|)^{-a-2r}(\prod_{i=1}^{r}\theta_i)\sin\langle n^{-1}j,\theta\rangle\,d\theta_1\ldots d\theta_d, \qquad (3.4.3)$$

where $c_1,\ldots,c_r>0$ and each summation has a finite number of terms which does not grow with n. The first integral is $(d-r)$-dimensional, the next $(d-r+1)$-dimensional, and the last one is d-dimensional. We have used the notation $\|\theta\|_{(-r)}$ and $\langle n^{-1}j,\theta\rangle_{(-r)}$ to indicate that r of the d variables θ_i have been replaced by the limits of integration as indicated in the two steps above for f and g.

As can be checked, each term in (3.4.3) is bounded by $c_i(n\delta)^{-a+d-r}$ for appropriate constants $c_i>0$, and thus it follows that

$$T\equiv\sum_{p=1}^{l}\rho_p=O\{(n\delta)^{d-a-r}\}.$$

This shows that (3.4.2) holds. Next observe that

$$\sum_{j\notin\mathcal{R}}\mathcal{B}(j)^2=\sum_{r=1}^{d}\sum_{j\in\mathcal{S}(r)}\mathcal{B}(j)^2. \qquad (3.4.4)$$

It now remains to calculate $\sum_{j\in\mathcal{S}(r)}\mathcal{B}(j)^2$. Using (3.4.2), one has for $1\le r\le d$

$$\begin{aligned}\sum_{j\in\mathcal{S}(r)}\mathcal{B}(j)^2 &\le c_1n^{2r}(n\delta)^{2(d-r-a)}\sum_{j\in\mathcal{S}(r)}(j_1^2)^{-r}\\ &\le c_2n^{2r}(n\delta)^{2(d-r-a)}n^{-r}\\ &= c_2n^{r}(n\delta)^{2(d-r-a)}\end{aligned}$$

for constants $c_1,c_2>0$. Substitution of the last estimate into (3.4.4) leads to

$$\begin{aligned}\sum_{j\notin\mathcal{R}}\mathcal{B}(j)^2 &\le c_3\max_{r=1,\ldots,d}n^{r}(n\delta)^{2(d-r-a)} \qquad (c_3>0)\\ &= O\{(n\delta)^{2(d-a)}\max_{r=1,\ldots,d}n^{r}(n\delta)^{-2r}\},\end{aligned}$$

which is the desired result. □

Lemma 3.4.4 *Put*

$$k_\Theta(j)=\pi^{-d}\prod_{i=1}^{d}\int_0^{\delta}\{1+2\lambda^2(1-\lambda^{-1})(1-\cos\theta_j)\}\cos(j_i\theta_i)\,d\theta_i \quad \textit{as } n\to\infty.$$

Then

$$\sum_{j\notin\mathcal{R}}k_\Theta(j)^2=o\{\sum_{l}k_\Theta(l)^2\} \quad \textit{as } n\to\infty.$$

Proof of Lemma 3.4.4

Let $j \in \mathbb{Z}^d \setminus \mathcal{R}$. By (3.2.13), there is $1 \le i \le d$ such that $|j_i| > Kn$. The function k_Θ is symmetric, and we may thus assume that for $j \notin \mathcal{R}$, $j = (j_1, \ldots, j_d)^T$, $|j_1| > Kn$. It follows that

$$\begin{aligned} \sum_{j \notin \mathcal{R}} k_\Theta(j)^2 &\le c_1 \sum_{|j_1|>Kn} \sum_{j_2 \in \mathbb{Z}} \cdots \sum_{j_d \in \mathbb{Z}} k_\Theta(j)^2 \\ &\le c_2 \{\sum_{|l|>n} e(l)^2\}\{\sum_{l \in \mathbb{Z}} e(l)^2\}^{d-1} \end{aligned} \tag{3.4.5}$$

for constants c_1, $c_2 > 0$. Here $e : \mathbb{Z} \to \mathbb{R}$ is given by

$$e(l) = \pi^{-1} \int_0^\delta \{1 + 2\lambda^2(1 - \lambda^{-1})(1 - \cos u)\} \cos(lu)\, du. \tag{3.4.6}$$

To prove the lemma, it suffices to show that

$$\sum_{|l|>n} e(l)^2 = o\{\sum_{l \in \mathbb{Z}} e(l)^2\}, \tag{3.4.7}$$

since

$$\sum_{l \in \mathbb{Z}^d} k_\Theta(l)^2 = \{\sum_{l \in \mathbb{Z}} e(l)^2\}^d.$$

Put $\mu = 2\lambda^2(1 - \lambda^{-1})$ and consider

$$\begin{aligned} \pi e(l) &= \int_0^\delta \{1 + \mu(1 - \cos u)\} \cos(lu)\, du \\ &= (1 + \mu) \int_0^\delta \cos(lu)\, du - \mu \int_0^\delta \cos u \cos(lu)\, du \\ &= (1 + \mu) l^{-1} \sin(l\delta) - \mu(l^2 - 1)^{-1}\{l \cos\delta \sin(l\delta) - \sin\delta \cos(l\delta)\}, \end{aligned}$$

where we have used the identity

$$\int_0^\delta \cos x \cos ax\, dx = (a^2 - 1)^{-1}\{a \cos\delta \sin(a\delta) - \sin\delta \cos(a\delta)\}.$$

Collecting terms in $\sin(l\delta)$, one gets

$$\pi e(l) = l^{-1} \sin(l\delta)(1 + \mu[1 - \cos\delta\{1 + (l^2 - 1)^{-1}\}]) + (l^2 - 1)^{-1} \mu \sin\delta \cos(l\delta)$$

and thus

$$\begin{aligned} e(l)^2 \le\ & c_3\{l^{-2} \sin^2(l\delta)(1 + \mu[1 - \cos\delta\{1 + (l^2 - 1)^{-1}\}])^2 \\ & + (l^2 - 1)^{-2} \mu^2 \sin^2\delta \cos^2(l\delta)\} \end{aligned}$$

for some $c_3 > 0$.

If $\lambda > 1$, then $\mu \leq 2\lambda^2$. It therefore follows that

$$\begin{aligned} e(l)^2 &\leq c_4\{l^{-2}\sin^2(l\delta)(1+2\lambda^2[1-\cos\delta\{1+(l^2-1)^{-1}\}])^2 \\ &\qquad +l^{-4}\lambda^4\sin^2\delta\cos^2(l\delta)\} \quad (c_4>0) \\ &\leq c_5(l^{-2}[1+\lambda^4\sin^4(\delta/2)\{1+o(n^{-1})\}]+l^{-4}\lambda^4\sin^2\delta) \quad (c_5>0), \end{aligned}$$

since $\sin^2(l\delta) \leq 1$, $\cos^2(l\delta) \leq 1$, and since $|l| > n$ implies that $(l^2-1)^{-1} = o(n^{-1})$. Next observe that for small δ, $\sin\delta \asymp \delta$, and thus

$$\begin{aligned} e(l)^2 &\leq c_6[l^{-2}\{1+(\lambda\delta)^4\}\{1+o(n^{-1})\}+l^{-4}\lambda^4\delta^2] \\ &\leq c_7 l^{-2}\{1+(\lambda\delta)^4+(\lambda\delta)^2\}\{1+o(n^{-1})\} \\ &\leq c_8 l^{-2}\max\{1,(\lambda\delta)^4\}\{1+o(n^{-1})\} \end{aligned} \tag{3.4.8}$$

for constants c_6, c_7, $c_8 > 0$. To obtain the second inequality in (3.4.8), we used the fact that $l^{-1}\lambda = O(1)$, which follows since $|l| > n$ and $\lambda = O(n)$.

To show (3.4.7), we now use the estimate (3.4.8) and obtain

$$\begin{aligned} \sum_{|l|>n} e(l)^2 &\leq c_8\max\{1,(\lambda\delta)^4\}\{1+o(n^{-1})\}\sum_{|l|>n} l^{-2} \\ &\leq c_9\max\{1,(\lambda\delta)^4\}n^{-1}\{1+o(n^{-1})\}. \end{aligned}$$

The desired result now follows, since $n\delta \to \infty$ implies that $n^{-1} = o(\delta)$. □

Lemma 3.4.5 *For $n > 0$ and $a > d$, put*

$$D_j = (2d)^{-1}\sum_{|k|=1}(h\star t)(j-k)-(h\star t)(j) \quad \textit{for } j \in \mathbb{Z}^d.$$

Then $D_j = o(1)$ for $j \in \mathcal{R}$, and thus $\sum_{j\in\mathcal{R}} D_j = o(n^d)$ as $n \to \infty$.

Proof of Lemma 3.4.5

For $j \in \mathcal{R}$,

$$\begin{aligned} |D_j| &= |(2d)^{-1}\sum_{|k|=1}\sum_{l}\{t(j-k-l)-t(j-l)\}h(l)| \\ &\leq \sum_l h(l)\sup_{|k|=1}|t(j-k-l)-t(j-l)|, \end{aligned}$$

since $h \geq 0$. Furthermore, for $j \in \mathbb{Z}^d$, $|k| = 1$,

$$\begin{aligned} |t(j-k)-t(j)| &\leq c_1\int_\Omega \tau(\theta)|e^{i\langle k,\theta\rangle}-1|\,d\theta \\ &\leq c_2\int_{n\Omega}(1+\|\theta\|)^{-a}|e^{i\langle k/n,\theta\rangle}-1|\,d\theta \end{aligned}$$

for $c_1, c_2 > 0$, where $n\Omega = \{n\theta \ : \ \theta \in \Omega\}$. Now, $|k| = 1$, and thus $k/n \to 0$ as $n \to \infty$. This implies that $|e^{i\langle k/n,\theta\rangle} - 1| \to 0$ as $n \to \infty$ for $\theta \in \Omega$, and therefore

$$|t(j-k) - t(j)| = o(1)$$

follows, since $\int_{n\Omega}(1 + \|\theta\|)^{-a} \, d\theta = O(1)$ by Lemma 3.4.1.

Since h is bounded and $\sum_l h(l) = 1$, the desired result now follows. $\square$

Lemma 3.4.6 *Put*

$$\begin{aligned} P &= (2\pi)^{-d} \int_{\Omega\setminus\Theta} \tau(\theta)^2 \{1 - \chi(\theta)\} \, d\theta, \\ Q &= (2\pi)^{-d} \int_{\Omega\setminus\Theta} \tau(\theta)^q \{1 - \chi(\theta)^{-1}\}^p \, d\theta, \quad \textit{for } p, q = 1, 2, \ldots. \end{aligned}$$

1. *If $a > d/2 + 1$, then, as $n \to \infty$,*

$$P \asymp n^d (n\delta)^{d-2a} \min\{1, (\lambda\delta)^2\}.$$

2. *If $\lambda\delta < 1$ and $a > (2p+1)d/q$, then, as $n \to \infty$,*

$$Q \asymp n^{(d-a)q} \delta^{d-aq} (\lambda\delta)^{2p}.$$

Proof of Lemma 3.4.6

To show part 1, take $\theta \in [0, \pi]$. Then $1 - \cos\theta \asymp \theta^2$, and therefore

$$\{1 - \chi(\theta)\} \asymp 1 - \prod_{i=1}^{d} \{1 + (\lambda\theta_i)^2\}^{-1} \asymp \min\{1, \lambda^2 \|\theta\|^2\}.$$

Put $\mathsf{C}\Theta = \Omega \setminus \Theta$. Using the above $\asymp$-relationship for χ, P is estimated by

$$\begin{aligned} P &\asymp n^{2d} \int_{\mathsf{C}\Theta} (1 + \|n\theta\|)^{-2a} \min(1, \lambda^2 \|\theta\|^2) \, d\theta \\ &\asymp n^{2d} \int_{\delta}^{\pi} (1 + n\rho)^{-2a} \min(1, \lambda^2 \rho^2) \rho^{d-1} \, d\rho, \end{aligned}$$

where $\rho = \|\theta\|$, and where we have used the fact that the Jacobian is proportional to ρ^{d-1}.

Next we distinguish two cases: $\lambda\delta < 1$ and $\lambda\delta \geq 1$.

If $\lambda\delta \geq 1$, then $\min(1, \lambda^2\rho^2) = 1$ and thus

$$P \asymp n^d \int_{n\delta}^{n\pi} (1 + \rho)^{-2a} \rho^{d-1} \, d\rho \asymp n^d (n\delta)^{d-2a}.$$

If $\lambda\delta < 1$, then

$$\begin{aligned}
P &\asymp n^{2d}\{\int_{\delta}^{\lambda^{-1}} (1+n\rho)^2(\lambda\rho)^2\rho^{d-1}\,d\rho + \int_{\lambda^{-1}}^{\pi}(1+n\rho)^{-2a}\rho^{d-1}\,d\rho\} \\
&\asymp n^d\{(\lambda n^{-1})^2\int_{n\delta}^{n\lambda^{-1}} (1+\rho)^{-2a}\rho^{d+1}\,d\rho + \int_{n\lambda^{-1}}^{n\pi}(1+\rho)^{-2a}\rho^{d-1}\,d\rho\} \\
&\asymp n^d(n\delta)^{d-2a}(\lambda\delta)^2.
\end{aligned}$$

Combining the two cases gives the desired result.

To show part 2, take $\theta \in [0,\pi]$. Then

$$\lambda^2\|\theta\|^2 \le \chi(\theta)^{-1} - 1 \le \lambda^2\|\theta\|^2 + R_\theta,$$

where

$$0 \le R_\theta \le c_1(\lambda\|\theta\|)^{2d}\,, \qquad c_1 > 0.$$

An upper bound for Q is

$$\begin{aligned}
Q &\le c_2 n^{qd}\int_{\mathsf{C}\Theta}(1+\|n\theta\|)^{-qa}\{(\lambda\|\theta\|)^{2p} + (\lambda\|\theta\|)^{2pd}\}\,d\theta \\
&\asymp n^{qd}\int_{\delta}^{\pi}(1+n\rho)^{-qa}\lambda^{2p}\rho^{2p+d-1}\{1+(\lambda\rho)^{2p(d-1)}\}\,d\rho \\
&\asymp n^{(q-1)d-2p}\int_{n\delta}^{n\pi}(1+\rho)^{-qa}\lambda^{2p}\rho^{2p+d-1}\{1+(\lambda n^{-1}\rho)^{2p(d-1)}\}\,d\rho \\
&\asymp n^{q(d-a)}\delta^{d-qa}(\lambda\delta)^{2p},
\end{aligned}$$

for some $c_2 > 0$, since $\lambda\delta < 1$ and $qa > (2p+1)d$.

Similarly, a lower bound for Q is obtained as

$$\begin{aligned}
Q &\ge c_3 n^{qd}\int_{\mathsf{C}\Theta}(1+\|n\theta\|)^{-qa}(\lambda\|\theta\|)^{2p}\,d\theta \\
&\asymp n^{q(d-a)}\delta^{d-qa}(\lambda\delta)^{2p},
\end{aligned}$$

and thus the required estimate for Q holds. □

3.4.2 Proof of Proposition 3.1

Proposition 3.1 *Assume that t, h and ϵ satisfy* **A1–A3** *and that $a > d/2$. If $\lambda\delta \to c$ where $0 \le c \le \infty$ as $n \to \infty$, then, as $n \to \infty$,*

$$MSSE \asymp n^d[(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}].$$

To show this proposition, we require Lemmas 3.4.1–3.4.3, which were stated and proved in the previous subsection.

Proof of Proposition 3.1

For $n > 0$, put

$$M_\infty = \sum_j \mathcal{B}(j)^2 + \sum_{j\in\mathcal{R}} \mathcal{V}(j), \tag{3.4.9}$$

where, the reader may recall, $\sum_j$ denotes $\sum_{j\in\mathbb{Z}^d}$. The difference between M_∞ and MSSE consists in the bias term; in MSSE, bias is restricted to $\mathcal{R}$, while bias in M_∞ is summed over $\mathbb{Z}^d$.

We first show that $\sum_j \mathcal{B}(j)^2 \asymp n^d(n\delta)^{d-2a}$, and thus

$$M_\infty \asymp n^d\{(n\delta)^{d-2a} + \delta^d \max(1, \lambda^4\delta^4)^d\}. \tag{3.4.10}$$

To prove the proposition, it then suffices to show that

$$\sum_{j\notin\mathcal{R}} \mathcal{B}(j)^2 = o\{\sum_j \mathcal{B}(j)^2\}. \tag{3.4.11}$$

Put $B_\infty = \sum_j \mathcal{B}(j)^2$. By Parseval's identity (see (1.3.9)),

$$B_\infty = (2\pi)^{-d} \int_{\Omega\setminus\Theta} \tau(\theta)^2 \, d\theta,$$

and thus by Lemma 3.4.1,

$$B_\infty \asymp n^d(n\delta)^{d-2a}.$$

Furthermore, from Lemma 3.4.2,

$$\sum_{j\in\mathcal{R}} \mathcal{V}(j) \asymp \delta^d \max(1, \lambda^4\delta^4)^d.$$

Together, these last two results show (3.4.10). As shown in Lemma 3.4.3,

$$\sum_{j\notin\mathcal{R}} \mathcal{B}(j)^2 = O\left[(n\delta)^{2(d-a)} \max_{r=1,\ldots,d}\{n^r(n\delta)^{-2r}\}\right]$$

and since $n\delta \to \infty$ and $r \le d$, it thus follows that

$$\sum_{j\notin\mathcal{R}} \mathcal{B}(j)^2 = o(B_\infty).$$

But $B_\infty = \sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 + \sum_{j\notin\mathcal{R}} \mathcal{B}(j)^2$, and it now follows that

$$\sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 \asymp B_\infty,$$

and therefore

$$\text{MSSE} \asymp M_\infty,$$

as required. □

3.4.3 Proof of Proposition 3.4

Proposition 3.4 *Assume that t, h and ϵ satisfy* **A1–A3**. *Assume further that $a > d/2$ and $\mathbf{E}\epsilon^4 < \infty$. If $\lambda\delta \to c$ where $0 \le c \le \infty$ as $n \to \infty$, then, as $n \to \infty$,*

$$\mathbf{E}(SSE - MSSE)^2 = o(MSSE^2).$$

The proof of Proposition 3.4 requires Lemmas 3.4.1–3.4.4, which were given in Subsection 3.4.1.

Proof of Proposition 3.4

Recall from (3.2.14) that SSE $= \sum_{j\in\mathcal{R}}\{\hat{t}(j) - t(j)\}^2$ and MSSE $= \mathbf{E}$(SSE). Since

$$\begin{aligned}
\hat{t}(j) &= t_\Theta(j) + (k_\Theta \star \epsilon)(j), \\
t_\Theta(j) &= (2\pi)^{-d} \int_\Theta \tau(\theta) e^{-i\langle j,\theta\rangle}\, d\theta, \\
k_\Theta(j) &= (2\pi)^{-d} \int_\Theta \chi(\theta)^{-1} e^{-i\langle j,\theta\rangle}\, d\theta,
\end{aligned}$$

(as in (3.3.2)), SSE and MSSE can be re-written in the following form.

$$\begin{aligned}
\text{SSE} &= \sum_{j\in\mathcal{R}} \{\mathcal{B}(j) + (k_\Theta \star \epsilon)(j)\}^2 \\
\text{MSSE} &= \sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 + n^d\sigma^2 \sum_j k_\Theta(j)^2,
\end{aligned} \tag{3.4.12}$$

where $\mathcal{B}(j)$ denotes the bias of $\hat{t}(j)$ which is given in (3.3.3), and $V = n^d\sigma^2\sum_j k_\Theta(j)^2$ denotes the variance of $\hat{t}$ on $\mathcal{R}$ as derived in (3.3.4).

Consider

$$\begin{aligned}
T &= \sum_{j\in\mathcal{R}} (k_\Theta \star \epsilon)(j)^2 - \sigma^2 n^d \sum_l k_\Theta(l)^2 \\
&= \sum_l \sum_i \epsilon_i\epsilon_l \sum_{j\in\mathcal{R}} k_\Theta(j-i)k_\Theta(j-l) - \sigma^2 \sum_{j\in\mathcal{R}} \sum_l k_\Theta(l)^2 \\
&= \sum_i (\epsilon_i^2 - \sigma^2) \sum_{j\in\mathcal{R}} k_\Theta(j-i)^2 + \sum_i \sum_{l\neq i} \epsilon_i\epsilon_l \sum_{j\in\mathcal{R}} k_\Theta(j-i)k_\Theta(j-l).
\end{aligned}$$

Put

$$\begin{aligned}
S &= \sum_{j\in\mathcal{R}} \mathcal{B}(j)(k_\Theta \star \epsilon)(j), \\
T_1 &= \sum_i (\epsilon_i^2 - \sigma^2) \sum_{j\in\mathcal{R}} k_\Theta(j-i)^2, \\
T_2 &= \sum_i \sum_{l\neq i} \epsilon_i\epsilon_l \sum_{j\in\mathcal{R}} k_\Theta(j-i)k_\Theta(j-l),
\end{aligned} \tag{3.4.13}$$

then

$$\begin{aligned}\text{SSE} - \text{MSSE} &= \sum_{j\in\mathcal{R}}\{2\mathcal{B}(j)(k_\Theta \star \epsilon)(j) + (k_\Theta \star \epsilon)(j)^2\} - \sigma^2 n^d \sum_j k_\Theta(j)^2 \\ &= 2S + T = 2S + T_1 + T_2, \end{aligned} \tag{3.4.14}$$

and $\mathbf{E}(\text{SSE} - \text{MSSE})^2 \leq c_1(\mathbf{E}S^2 + \mathbf{E}T_1^2 + \mathbf{E}T_2^2)$, for some constant $c_1 > 0$. It thus suffices to show that

$$\mathbf{E}P^2 = o(\text{MSSE}^2) \qquad \text{for } P = S, T_1, T_2. \tag{3.4.15}$$

Consider

$$\begin{aligned}\mathbf{E}S^2 &= \mathbf{E}\{\sum_{j\in\mathcal{R}} \mathcal{B}(j)(k_\Theta \star \epsilon)(j)\}^2 \\ &= \sum_{j\in\mathcal{R}}\sum_{l\in\mathcal{R}} \mathcal{B}(j)\mathcal{B}(l)\sum_i \mathbf{E}\epsilon_i^2 k_\Theta(j-i)k_\Theta(l-i) \\ &= \sigma^2 \sum_i\{\sum_{j\in\mathcal{R}} \mathcal{B}(j)k_\Theta(j-i)\}^2. \end{aligned} \tag{3.4.16}$$

Observe that the Fourier transform of $\mathcal{B}$ has support $\Omega \setminus \Theta$ while the Fourier transform of k_Θ has support Θ. From this it follows that

$$\sum_j \mathcal{B}(j)k_\Theta(j-l) = (2\pi)^{-d}\int_\Omega \tau(\theta)\mathcal{I}_{\Omega\setminus\Theta}\chi(\theta)^{-1}\mathcal{I}_\Theta e^{-i\langle l,\theta\rangle}\, d\theta = 0 \qquad \text{for } l \in \mathbb{Z}^d,$$

and therefore

$$\sum_{j\in\mathcal{R}} \mathcal{B}(j)k_\Theta(j-l) = -\sum_{j\notin\mathcal{R}} \mathcal{B}(j)k_\Theta(j-l). \tag{3.4.17}$$

To estimate (3.4.16), let $\mathcal{K}$ denote a subset of $\mathbb{Z}^d$. Using (3.4.17), one obtains

$$\begin{aligned}\sum_l\{\sum_{j\in\mathcal{R}} \mathcal{B}(j)k_\Theta(j-l)\}^2 &= \sum_{l\notin\mathcal{K}}\{\sum_{j\in\mathcal{R}} \mathcal{B}(j)k_\Theta(j-l)\}^2 + \sum_{l\in\mathcal{K}}\{\sum_{j\notin\mathcal{R}} \mathcal{B}(j)k_\Theta(j-l)\}^2 \\ &\leq \sum_{l\notin\mathcal{K}}\sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 \sum_{j\in\mathcal{R}} k_\Theta(j-l)^2 + \sum_{l\in\mathcal{K}}\sum_{j\notin\mathcal{R}} \mathcal{B}(j)^2 \sum_{j\notin\mathcal{R}} k_\Theta(j-l)^2 \\ &= \sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 \sum_{l\notin\mathcal{K}}\sum_{j\in\mathcal{R}} k_\Theta(j-l)^2 + \sum_{j\notin\mathcal{R}} \mathcal{B}(j)^2 \sum_{l\in\mathcal{K}}\sum_{j\notin\mathcal{R}} k_\Theta(j-l)^2. \end{aligned} \tag{3.4.18}$$

We have used Hölder's inequality here. Next observe that

$$\sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 \sum_{l\notin\mathcal{K}}\sum_{j\in\mathcal{R}} k_\Theta(j-l)^2 = o(\text{MSSE}^2), \tag{3.4.19}$$

provided that

$$\sum_{l\notin\mathcal{K}}\sum_{j\in\mathcal{R}} k_\Theta(j-l)^2 = o(\text{MSSE}), \tag{3.4.20}$$

since $\sum_{j\in\mathcal{R}} \mathcal{B}(j)^2 \leq \text{MSSE}$ by Proposition 3.1.

For the term $\sum_{j \notin \mathcal{R}} \mathcal{B}(j)^2 \sum_{l \in \mathcal{K}} \sum_{j \notin \mathcal{R}} k_\Theta(j-l)^2$ in (3.4.18), one may take $\mathcal{K} = k_1 \mathcal{R}$ for some $k_1 > 0$. In this case,

$$\sum_{l \in \mathcal{K}} \sum_{j \notin \mathcal{R}} k_\Theta(j-l)^2 = o(\mathrm{MSSE})$$

follows immediately from (3.4.20). Now $\sum_{j \notin \mathcal{R}} \mathcal{B}(j)^2 = o\{\sum_{j \in \mathcal{R}} \mathcal{B}(j)^2\}$ is a consequence of Lemma 3.4.3. The proof that $\mathbf{E}S^2 = o(\mathrm{MSSE}^2)$ thus reduces to showing (3.4.20). Since $\mathcal{K}$ in (3.4.20) is arbitrary, we may take $\mathcal{K} = \mathcal{R}$. Then

$$\sum_{j \in \mathcal{R}} \sum_{l \notin \mathcal{R}} k_\Theta(j-l)^2 \leq c_3 n^d \max_{j \in \mathcal{R}} \sum_{l \notin \mathcal{R}} k_\Theta(j-l)^2 \qquad (c_3 > 0),$$

and it thus suffices to show that

$$\max_{j \in \mathcal{R}} \sum_{l \notin \mathcal{R}} k_\Theta(j-l)^2 = o\{\sum_j k_\Theta(j)^2\}. \tag{3.4.21}$$

(Recall that $\sum_j k_\Theta(j)^2 = (2\pi)^{-d} \int_\Theta \chi(\theta)^{-2} d\theta \asymp \delta^d \max\{1, (\lambda\delta)^{4d}\}$ by (3.3.4) and Lemma 3.4.2, and that $n^d \sum_j k_\Theta(j)^2 \leq$ MSSE.) Equation (3.4.21) follows therefore from Lemma 3.4.4, and thus

$$\mathbf{E}S^2 = o(\mathrm{MSSE}^2)$$

as required.

Next consider

$$T_1 = \sum_i (\epsilon_i^2 - \sigma^2) \sum_{j \in \mathcal{R}} k_\Theta(j-i)^2$$

as given in (3.4.13). Observe that

$$\mathbf{E}\{\sum_i (\epsilon_i^2 - \sigma^2)\}^2 = \sum_i (\mathbf{E}\epsilon_i^4 - \sigma^4),$$

and thus it follows that

$$\begin{aligned} \mathbf{E}T_1^2 &= \sum_i (\mathbf{E}\epsilon_i^4 - \sigma^4) \sum_{j \in \mathcal{R}} \sum_{l \in \mathcal{R}} k_\Theta(j-i)^2 k_\Theta(j-l)^2 \\ &= \rho \sum_i Z(i,i)^2, \end{aligned} \tag{3.4.22}$$

where

$$\begin{aligned} \rho &= \mathbf{E}\epsilon^4 - \sigma^4 \\ Z(i,l) &= \sum_{j \in \mathcal{R}} k_\Theta(j-i) k_\Theta(j-l). \end{aligned} \tag{3.4.23}$$

Consider

$$\sum_i Z(i,i)^2 = \sum_i \sum_{j \in \mathcal{R}} \sum_{l \in \mathcal{R}} k_\Theta(j-i)^2 k_\Theta(l-i)^2$$

$$
\begin{aligned}
&\leq \sum_{j\in\mathcal{R}}\sum_i k_\Theta(j-i)^2\sum_l k_\Theta(l-i)^2 \\
&= \sum_{j\in\mathcal{R}}\{\sum_i k_\Theta(i)^2\}^2 \\
&= \sum_{j\in\mathcal{R}}\{(2\pi)^{-d}\int_\Theta \chi(\theta)^{-2}\, d\theta\}^2,
\end{aligned}
$$

where the last equality follows from Parseval's identity. By Lemma 3.4.2 applied to $\{\int_\Theta \chi(\theta)^{-2}\, d\theta\}^2$, we now have that

$$
\sum_i Z(i,i)^2 \leq c_5 n^d \delta^{2d} \max\{1,(\lambda\delta)^{8d}\} \leq c_6 n^{-d}\mathrm{MSSE}^2 = o(\mathrm{MSSE}^2) \qquad (3.4.24)
$$

for constants $c_5,\ c_6 > 0$. This shows that $\mathbf{E}T_1^2 = o(\mathrm{MSSE}^2)$ by substituting (3.4.24) into (3.4.22).

It remains to estimate $\mathbf{E}T_2^2$. By (3.4.13) and (3.4.23),

$$
T_2 = \sum_i \sum_{l\neq i} \epsilon_i \epsilon_l Z(i,l),
$$

and thus

$$
\begin{aligned}
\mathbf{E}T_2^2 &= \sum_i\sum_{l\neq i}\sum_m\sum_{j\neq m} \mathbf{E}(\epsilon_i\epsilon_j\epsilon_l\epsilon_m)Z(i,l)Z(m,j) \\
&= 2\sum_i\sum_{l\neq i}\sigma^4 Z(i,l)^2 \qquad (\text{since } Z(i,l)=Z(l,i)) \\
&= 2\sigma^4\sum_{m\in\mathcal{R}}\sum_{j\in\mathcal{R}}\sum_i\sum_{l\neq i} k_\Theta(m-i)k_\Theta(j-i)k_\Theta(m-l)k_\Theta(j-l) \\
&\leq c_7\sum_{m\in\mathcal{R}}\sum_{j\in\mathcal{R}}\{\sum_i k_\Theta(m-i)k_\Theta(j-i)\}^2 \qquad (c_7>0) \\
&= c_7\sum_{m\in\mathcal{R}}\sum_{j\in\mathcal{R}}\{(k_\Theta \star k_\Theta)(m-j)\}^2 \qquad (\text{by the symmetry of } k_\Theta) \\
&\leq c_7\sum_{m\in\mathcal{R}}\sum_j\{(k_\Theta \star k_\Theta)(j)\}^2 \\
&= c_8 n^d\int_\Theta \chi(\theta)^{-4}\, d\theta. \qquad (3.4.25)
\end{aligned}
$$

The last equality follows from Parseval's identity. A further application of Lemma 3.4.2 yields

$$
\mathbf{E}T_2^2 \leq c_9 n^d\delta^d \max\{1,(\lambda\delta)^{8d}\} \quad (c_9>0).
$$

Using the result of Proposition 3.1, it now follows that

$$
\mathbf{E}T_2^2 \leq c_{10}\mathrm{MSSE} = o(\mathrm{MSSE}^2) \quad (c_{10}>0),
$$

as required. This completes the proof of Proposition 3.4. □

3.4.4 Proof of Theorem 3.5

Theorem 3.5 *Assume that* t, h *and* ϵ *satisfy* **A1–A3**. *Assume further that* $a > 5d/2$, $\lambda > 1$ *and* $\mathbf{E}\epsilon^4 < \infty$. *If* δ_m *minimises MSSE over* δ *for given* n *and* λ, *then 1 and 2 below are equivalent:*

1. $\mathbf{E}\{(CV_1 - SSE)(\delta_m)^2\} = o\{MSSE(\delta_m)^2\}$;
2. $\lambda = o(n^{1-d/2a})$.

In the proof of this theorem, Lemmas 3.4.1–3.4.3, 3.4.5–3.4.6 are required. These together with their proofs were given in Subsection 3.4.1.

Proof of Theorem 3.5

Fix $n > 0$, and write $\mathcal{R}$ for $\mathcal{R}_n$. Recall from (3.3.8)–(3.3.10) that

$$\mathrm{CV}_1 = \mathrm{SSE}_1^* + N_1,$$

where

$$\mathrm{SSE}_1^* = \sum_{j\in\mathcal{R}} \{\hat{t}(j)^2 - 2\hat{t}_1^*(j)\tilde{t}_1(j) + t(j)^2\}$$

is the first cross-validation approximation to SSE, and

$$N_1 = 2\sum_{j\in\mathcal{R}} [t(j)\{\tilde{t}_1(j) - t(j)\} + \{(h\star t)(j) - t(j)\}(\epsilon\star h^{(-1)})(j)]$$

contains only those terms which do not depend on δ. Thus N_1 is omitted when comparing SSE_1^* and SSE.

For notational simplicity we now omit the subscript 1 and write CV, SSE^*, $\hat{t}^*$ and $\tilde{t}$ instead of CV_1, SSE_1^*, $\hat{t}_1^*$ and $\tilde{t}_1$ in this proof. Put

$$\begin{aligned} U &= \sum_{j\in\mathcal{R}} \{\hat{t}^*(j) - \hat{t}(j)\}(h\star t)(j), \\ V &= \sum_{j\in\mathcal{R}} \{\hat{t}^*(j) - t(j)\}\epsilon_j, \quad \text{and} \\ W &= \sum_{j\in\mathcal{R}} \{\hat{t}(j) - t(j) - (\epsilon\star h^{(-1)})(j)\}\{(h\star t)(j) - t(j)\}. \end{aligned} \tag{3.4.26}$$

We have

$$\mathrm{CV} - \mathrm{SSE} = -2(U + V + W).$$

To prove the theorem, we show the following

- $\mathbf{E}(U^2 + V^2) = o(\mathrm{MSSE}^2)$;

- $\mathbf{E}\{W(\delta_m)^2\} = o\{\mathrm{MSSE}(\delta_m)^2\}$ if and only if $\lambda = o(n^{1-d/2a})$,

where δ_m denotes the minimiser of MSSE, $W(\delta_m)$ is the value of W at δ_m, and

$$\mathrm{MSSE} \asymp n^d\{(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}$$

by Proposition 3.1.

Define an operator D on points y_j, $j \in \mathbb{Z}^d$ by

$$\mathsf{D}y_j = (2d)^{-1} \sum_{|l|=1} y_{j+l} - y_j. \tag{3.4.27}$$

Consider

$$\begin{aligned} U &= \sum_{j\in\mathcal{R}} \{\hat{t}^*(j) - \hat{t}(j)\}(h\star t)(j) \\ &= \sum_{j\in\mathcal{R}} \{(h\star t)(j)\mathsf{D}(h\star t)(j) + (h\star t)(j)\mathsf{D}\epsilon_j\}k_\Theta(0). \end{aligned} \tag{3.4.28}$$

Now,

$$\sup_{j\in\mathcal{R}} |(h\star t)(j)| \le \sup_{j\in\mathcal{R}} |t(j)| \le (2\pi)^{-d} \int_\Omega \tau(\theta)\, d\theta = O(1) \tag{3.4.29}$$

by Lemma 3.4.1.

For the first term in (3.4.28), which we denote by U_1, we have the following estimate:

$$\begin{aligned} |U_1| &= |\sum_{j\in\mathcal{R}} (h\star t)(j)\mathsf{D}(h\star t)(j)k_\Theta(0)| \\ &\le |k_\Theta(0)| \sup_{j\in\mathcal{R}} |(h\star t)(j)| \sum_{j\in\mathcal{R}} |\mathsf{D}(h\star t)(j)| \\ &= o[n^d\delta^d \max\{1, (\lambda\delta)^{2d}\}]. \end{aligned} \tag{3.4.30}$$

Here we have used the fact that $k_\Theta(0) = O[\delta^d \max\{1, (\lambda\delta)^{2d}\}]$ and that $\sum_{j\in\mathcal{R}} \mathsf{D}(h\star t)(j) = o(n^d)$ which follow from Lemmas 3.4.2 and 3.4.5.

For the random term in (3.4.28) observe that

$$\mathbf{E}(\sum_{j\in\mathcal{R}} \mathsf{D}\epsilon_j)^2 = O(n^d),$$

since the ϵ's are independent. Using (3.4.29) and Lemma 3.4.2, one therefore obtains for

$$U_2 \equiv \sum_{j\in\mathcal{R}} (h\star t)(j)\mathsf{D}\epsilon_j k_\Theta(0)$$

the following estimate:

$$\begin{aligned}\mathbf{E}U_2^2 &\le |k_\Theta(0)|^2 \sup_{j\in\mathcal{R}} |(h\star t)(j)|^2 \mathbf{E}(\sum_{j\in\mathcal{R}} \mathsf{D}\epsilon_j)^2 \\ &= O[n^d\delta^{2d}\max\{1,(\lambda\delta)^{4d}\}]. \end{aligned} \tag{3.4.31}$$

Combining (3.4.30) and (3.4.31) therefore yields

$$\begin{aligned}\mathbf{E}U^2 &\le c_1(U_1^2+\mathbf{E}U_2^2) \quad (c_1>0) \\ &= o[(n\delta)^{2d}\max\{1,(\lambda\delta)^{4d}\}] = o(\text{MSSE}^2). \end{aligned} \tag{3.4.32}$$

We next estimate $\mathbf{E}V^2$, V as in (3.4.26). From (3.2.18) and (3.3.2) it follows that the estimator $\hat{t}^*(j)$ is

$$\begin{aligned}\hat{t}^*(j) &= \hat{t}(j)+k_\Theta(0)\mathsf{D}X_j \\ &= t_\Theta(j)+k_\Theta(0)\mathsf{D}(h\star t)(j)+(k_\Theta\star\epsilon)(j)+k_\Theta(0)\mathsf{D}\epsilon_j. \end{aligned}$$

Using this equality, V becomes

$$\begin{aligned}V &= \sum_{j\in\mathcal{R}}\{\hat{t}^*(j)-t(j)\}\epsilon_j \\ &= \sum_{j\in\mathcal{R}}[\{\mathcal{B}(j)+k_\Theta(0)\mathsf{D}(h\star t)(j)\}\epsilon_j+\{(k_\Theta\star\epsilon)(j)+k_\Theta(0)\mathsf{D}\epsilon_j\}\epsilon_j], \end{aligned} \tag{3.4.33}$$

where $\mathcal{B}(j)$ denotes the bias of $\hat{t}(j)$ (see (3.3.3)). Put

$$V_1 = \sum_{j\in\mathcal{R}}\{\mathcal{B}(j)+k_\Theta(0)\mathsf{D}(h\star t)(j)\}\epsilon_j. \tag{3.4.34}$$

Then

$$\begin{aligned}\mathbf{E}V_1^2 &= \sigma^2\sum_{j\in\mathcal{R}}\{\mathcal{B}(j)+k_\Theta(0)\mathsf{D}(h\star t)(j)\}^2 \\ &\le c_2[\sum_j \mathcal{B}(j)^2+k_\Theta(0)^2\sum_{j\in\mathcal{R}}\{\mathsf{D}(h\star t)(j)\}^2] \end{aligned} \tag{3.4.35}$$

for some $c_2>0$. Now,

$$\sum_j \mathcal{B}(j)^2 = (2\pi)^{-d}\int_{\Omega\setminus\Theta}\tau(\theta)^2\,d\theta = O\{n^d(n\delta)^{d-2a}\}$$

follows from Parseval's identity and Lemma 3.4.1. Furthermore

$$\begin{aligned}k_\Theta(0)^2 &= O\{\delta^{2d}\max\{1,(\lambda\delta)^{4d}\}, \\ \sum_{j\in\mathcal{R}}\{\mathsf{D}(h\star t)(j)\}^2 &= o(n^d) \end{aligned}$$

follow from Lemmas 3.4.2 and 3.4.5. The term $\mathbf{E}V_1^2$ of (3.4.35) is therefore bounded by

$$\begin{aligned}\mathbf{E}V_1^2 &= O\{n^d(n\delta)^{d-2a}\} + o[n^d\delta^{2d}\max\{1,(\lambda\delta)^{4d}\}] \\ &= o(\mathrm{MSSE}^2). \end{aligned} \tag{3.4.36}$$

Let V_2 denote the following sum of terms in (3.4.33)

$$V_2 = \sum_{j\in\mathcal{R}} \epsilon_j\{(k_\Theta \star \epsilon)(j) + k_\Theta(0)\mathsf{D}\epsilon_j\}. \tag{3.4.37}$$

Observe that

$$\begin{aligned} k_\Theta(0)\mathsf{D}\epsilon_j + (k_\Theta\star\epsilon)(j) &= k_\Theta(0)\{(2d)^{-1}\sum_{|l|=1}\epsilon_{j-l} - \epsilon_j\} + \sum_l \epsilon_{j-l}k_\Theta(l) \\ &= k_\Theta(0)(2d)^{-1}\sum_{|l|=1}\epsilon_{j-l} + \sum_{l\neq 0}\epsilon_{j-l}k_\Theta(l), \end{aligned}$$

and therefore

$$V_2 = \sum_{j\in\mathcal{R}}\epsilon_j\{(2d)^{-1}\sum_{|l|=1}\epsilon_{j-l}k_\Theta(0) + \sum_{l\neq 0}\epsilon_{j-l}k_\Theta(l)\}. \tag{3.4.38}$$

We now bound $\mathbf{E}V_2^2$ in the following way:

$$\begin{aligned} \mathbf{E}V_2^2 &= \sum_{j_1\in\mathcal{R}}\sum_{j_2\in\mathcal{R}}\mathbf{E}\{(2d)^{-2}\sum_{|l_1|=1}\sum_{|l_2|=1}\epsilon_{j_1}\epsilon_{j_2}\epsilon_{j_1-l_1}\epsilon_{j_2-l_2}k_\Theta(0)^2 \\ &\qquad + d^{-1}\sum_{|l_1|=1}\sum_{l_2\neq 0}\epsilon_{j_1}\epsilon_{j_2}\epsilon_{j_1-l_1}\epsilon_{j_2-l_2}k_\Theta(0)k_\Theta(l_2) \\ &\qquad + \sum_{l_1\neq 0}\sum_{l_2\neq 0}\epsilon_{j_1}\epsilon_{j_2}\epsilon_{j_1-l_1}\epsilon_{j_2-l_2}k_\Theta(l_1)k_\Theta(l_2)\} \\ &\leq c_3\{\sum_{j\in\mathcal{R}}k_\Theta(0)^2 + \sum_{j\in\mathcal{R}}\sum_{|l_1|=1}k_\Theta(0)k_\Theta(l) + \sum_{j\in\mathcal{R}}\sum_l k_\Theta(l)^2\} \\ &\leq c_4 n^d\{k_\Theta(0)^2 + \sum_l k_\Theta(l)^2\}, \end{aligned}$$

where $c_3,\ c_4 > 0$ and

$$\sum_{|l|=1}k_\Theta(0)k_\Theta(l) \leq c_5 k_\Theta(0)\{\sum_{|l|=1}k_\Theta(l)^2\}^{1/2} \leq c_6 k_\Theta(0)^2$$

was used for some $c_5,\ c_6 > 0$. Writing $k_\Theta(0)$ in terms of its inverse Fourier transform and applying Parseval's identity to $\sum_l k_\Theta(l)^2$ yields, for $c_7 > 0$,

$$\begin{aligned} \mathbf{E}V_2^2 &\leq c_7 n^d[(2\pi)^{-2d}\{\int_\Theta \chi(\theta)^{-1}\,d\theta\}^2 + (2\pi)^{-d}\int_\Theta \chi(\theta)^{-2}\,d\theta] \\ &= O[(n\delta)^d\max\{1,(\lambda\delta)^{4d}\}] = o(\mathrm{MSSE}^2) \end{aligned} \tag{3.4.39}$$

by Lemma 3.4.2.

Combining (3.4.32), (3.4.36) and (3.4.39), one obtains that

$$\mathbf{E}(U^2+V^2)=o(\mathrm{MSSE}^2).$$

It remains to show that if δ_m minimises MSSE over δ, then $\mathbf{E}\{W(\delta_m)^2\}=o\{\mathrm{MSSE}(\delta_m)^2\}$ if and only if $\lambda=o(n^{1-d/2a})$, where

$$W(\delta_m)=\sum_{j\in\mathcal{R}}\{\hat{t}_{\delta_m}(j)-t(j)-(\epsilon\star h^{(-1)})(j)\}\{(h\star t)(j)-t(j)\}.$$

Put

$$\begin{aligned} r(j) &= k_{\Omega\setminus\Theta}(j)=(2\pi)^{-d}\int_{\Omega\setminus\Theta}\chi(\theta)^{-1}e^{-i\langle j,\theta\rangle}\,d\theta,\\ s(j) &= (h\star t)(j)-t(j). \end{aligned} \tag{3.4.40}$$

For $\mathcal{B}(j)$ denoting bias of $\hat{t}(j)$,

$$W=\sum_{j\in\mathcal{R}}\{\mathcal{B}(j)-(\epsilon\star r)(j)\}s(j), \tag{3.4.41}$$

and

$$\mathbf{E}W^2=\{\sum_{j\in\mathcal{R}}\mathcal{B}(j)s(j)\}^2+\sigma^2\sum_k\{\sum_{j\in\mathcal{R}}s(j)r(j-k)\}^2, \tag{3.4.42}$$

since

$$\begin{aligned} \mathbf{E}\{\sum_{j\in\mathcal{R}}(\epsilon\star r)(j)s(j)\}^2 &= \sigma^2\sum_{j_1\in\mathcal{R}}\sum_{j_2\in\mathcal{R}}s(j_1)s(j_2)\sum_k r(j_1-k)r(j_2-k)\\ &= \sigma^2\sum_k\{\sum_{j\in\mathcal{R}}s(j)r(j-k)\}^2. \end{aligned} \tag{3.4.43}$$

Instead of estimating $\mathbf{E}W^2$ directly, we first estimate the expression W_∞ given below in (3.4.45), in which the sums over $\mathcal{R}$ in $\mathbf{E}W^2$ are replaced by infinite sums over $\mathbb{Z}^d$, and then give reasons why this change can be made without affecting the asymptotic results.

Put

$$\begin{aligned} P &= (2\pi)^{-d}\int_{\Omega\setminus\Theta}\tau(\theta)^2\{\chi(\theta)-1\}\,d\theta,\\ Q &= (2\pi)^{-d}\int_{\Omega\setminus\Theta}\tau(\theta)^2\{\chi(\theta)^{-1}-1\}^2\,d\theta \end{aligned} \tag{3.4.44}$$

and consider

$$\begin{aligned} W_\infty &\equiv \{\sum_j\mathcal{B}(j)s(j)\}^2+\sigma^2\sum_k\{\sum_j s(j)r(j-k)\}^2\\ &= [(2\pi)^{-d}\int_{\Omega\setminus\Theta}\tau(\theta)^2\{\chi(\theta)-1\}\,d\theta]^2+\sigma^2(2\pi)^{-d}\int_{\Omega\setminus\Theta}\tau(\theta)^2\{\chi(\theta)^{-1}-1\}^2\,d\theta \end{aligned}$$

$$= \quad P^2 + \sigma^2 Q. \tag{3.4.45}$$

By Lemma 3.4.6,

$$P \asymp n^d (n\delta)^{d-2a} \min\{1, (\lambda\delta)^2\}, \tag{3.4.46}$$

and therefore

$$P/\text{MSSE} \asymp (n\delta)^{d-2a} \min\{1, (\lambda\delta)^2\}[(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}]^{-1}$$

follows from Proposition 3.1. To show that $P/\text{MSSE} \to 0$, let δ_m denote the minimiser of MSSE over δ. Then for $\delta = c_8\delta_m$, $c_8 > 0$

$$(P/\text{MSSE})(\delta) \asymp \min\{1, (\lambda\delta)^2\}. \tag{3.4.47}$$

From (3.4.47) it follows that

$$(P/\text{MSSE})(\delta) \to 0 \quad \text{if and only if} \quad \lambda\delta \to 0,$$

and this is equivalent to $\lambda = o(\delta^{-1})$. But $\delta_m = c_9 n^{-1+d/2a}$, $c_9 > 0$, is a consequence of Corollary 3.2, and therefore

$$P(\delta)^2 = o\{\text{MSSE}(\delta)^2\} \text{ if and only if } \lambda = o(n^{1-d/2a}). \tag{3.4.48}$$

To show that $Q = o(\text{MSSE}^2)$, in view of (3.4.48) it now suffices to assume that $\lambda\delta < 1$. Here we require the assumption that $a > 5d/2$ in order to show in Lemma 3.4.6 that

$$Q \asymp n^d (n\delta)^{d-2a} (\lambda\delta)^4. \tag{3.4.49}$$

Thus $Q = o(\text{MSSE}^2)$ follows, since $(\lambda\delta)^4 = o\{(n\delta)^d\}$. (By assumption $\lambda\delta \to 0$ and $n\delta \to \infty$.)

So far we have shown that for $\delta = c\delta_m$, $W_\infty = o(\text{MSSE}^2)$. It now remains to show that $\mathbf{E}W^2 = o(\text{MSSE}^2)$. Applying arguments similar to those given in the proof of Lemma 3.4.3 shows that

$$\sum_{j \notin \mathcal{R}} \mathcal{B}(j)s(j) = O[(n\delta)^{2(d-a)} \min\{1, (\lambda\delta)^2\} \max_{r=1,\ldots,d}\{n^r (n\delta)^{-2r}\}] = o(P)$$

(see also (3.4.46)); and

$$\sum_k \{\sum_{j \notin \mathcal{R}} s(j) r(j-k)\}^2 = o(\text{MSSE}^2).$$

From these last estimates the desired result follows immediately for $\mathbf{E}W^2$. □

3.4.5 Proof of Theorem 3.6

Theorem 3.6 *Assume that t, h and ϵ satisfy* **A1–A3**. *Assume further that $a > 5d/2$, $\lambda > 1$ and $\mathbf{E}\epsilon^4 < \infty$. Then 1 and 2 below are equivalent:*

1. $\mathbf{E}(CV_2 - SSE)^2 = o(MSSE^2)$;
2. $(n^{-1}\lambda^4\delta^2)^d\{1 + (\lambda\delta)^{4d}\}[(n\delta)^{d-2a} + \delta^d\{1 + (\lambda\delta)^{4d}\}]^{-2} \to 0$.

If δ_m minimises MSSE over δ for given n and λ, and if $\delta = c_3\delta_m$ for $c_3 > 0$, then 2 holds if and only if $\lambda = o(n^{1/4})$.

The proof of Theorem 3.6 requires Lemmas 3.4.1–3.4.3, 3.4.5–3.4.6 as well as

Lemma 3.4.7 *For $n > 0$ put*

$$B = \sum_{j\in\mathcal{R}}(\epsilon \star h^{(-1)})(j)\{(\epsilon \star k_\Theta)(j) + c_{\mathcal{M}}^{-1}\sum_{i\in\mathcal{N}} k_\Theta(i)(\sum_{l\in\mathcal{M}} \epsilon_{j-l} - \epsilon_{j-i})\},$$

where $c_{\mathcal{M}} = |\mathcal{M}|$. Then

$$\mathbf{E}B^2 = O[(n\lambda^4\delta^2)^d \max\{1, (\lambda\delta)^{4d}] + o(MSSE^2), \quad \textit{as } n \to \infty.$$

Proof of Lemma 3.4.7

Put $\kappa = c_{\mathcal{M}}^{-1}$. For $j \in \mathbb{Z}^d$,

$$\begin{aligned}(\epsilon \star k_\Theta)(j) \quad + \quad & \kappa\sum_{i\in\mathcal{N}} k_\Theta(j)(\sum_{l\in\mathcal{M}} \epsilon_{k-l} - \epsilon_{j-i}) \\ = \; & \sum_{l\notin j+\mathcal{N}} \epsilon_l k_\Theta(j-l) + \kappa \sum_{l\in j+\mathcal{M}} \epsilon_l \sum_{i\in j+\mathcal{N}} k_\Theta(j-i) \\ = \; & \sum_l \alpha_{jl}\epsilon_l,\end{aligned}$$

where

$$\alpha_{jl} = \begin{cases} k_\Theta(j-l) & \text{if } \ j-l \in \widetilde{\mathcal{M}} \cap \widetilde{\mathcal{N}} \\ k_\Theta(j-l) + \kappa\sum_{i\in j+\mathcal{N}} k_\Theta(j-i) & \text{if } \ j-l \in \mathcal{M} \\ 0 & \text{if } \ j-l \in \mathcal{N}, \end{cases} \tag{3.4.50}$$

and $\widetilde{\mathcal{M}} = \mathbb{Z}^d \setminus \mathcal{M}$, $\widetilde{\mathcal{N}} = \mathbb{Z}^d \setminus \mathcal{N}$. With this notation, B becomes

$$B = \sum_{j\in\mathcal{R}}\sum_m \epsilon_m h^{(-1)}(j-m)\sum_l \alpha_{jl}\epsilon_l = \sum_m\sum_l \epsilon_m\epsilon_l\beta_{ml},$$

where

$$\beta_{lm} = \sum_{j\in\mathcal{R}} \alpha_{jm} h^{(-1)}(j-l).$$

Note that $\beta_{ll} = 0$, since $\alpha_{jl} = 0$ for $j \in l + \mathcal{N}$, and $h^{(-1)}(j) = 0$ for $j \in \widetilde{\mathcal{N}}$.

Now,

$$\begin{aligned}
\mathbf{E}B^2 &= \sum_{l_1}\sum_{l_2}\sum_{k_1}\sum_{k_2} \mathbf{E}(\epsilon_{l_1}\epsilon_{l_2}\epsilon_{k_1}\epsilon_{k_2})\beta_{l_1k_1}\beta_{l_2k_2} \\
&= \sigma^4 \sum_l \sum_k (\beta_{kl}^2 + \beta_{lk}\beta_{kl}) \\
&= \sigma^4 \sum_{j_1\in\mathcal{R}} \sum_{j_2\in\mathcal{R}} [(\sum_l \alpha_{j_1l}\alpha_{j_2l})\{\sum_k h^{(-1)}(j_1-k)h^{(-1)}(j_2-k)\} \\
&\quad + \{\sum_l \alpha_{j_1l} h^{(-1)}(j_2-l)\}\{\sum_k \alpha_{j_2k} h^{(-1)}(j_1-k)\}].
\end{aligned}$$

Let γ denote the Fourier transform of α. An application of Parseval's identity to the sums over k and l yields

$$\begin{aligned}
\mathbf{E}B^2 &= \sigma^4(2\pi)^{-2d} \sum_{j_1\in\mathcal{R}} \sum_{j_2\in\mathcal{R}} [\{\int_\Omega \gamma(\theta)^2 e^{i\langle j_1-j_2,\theta\rangle}\, d\theta\}\{\int_\Omega \chi(\theta)^{-2} e^{i\langle j_1-j_2,\theta\rangle}\, d\theta\} \\
&\quad + \{\int_\Omega \gamma(\theta)\chi(\theta)^{-1} e^{i\langle j_1-j_2,\theta\rangle}\, d\theta\}^2]. \qquad (3.4.51)
\end{aligned}$$

Consider

$$\begin{aligned}
J &= (2\pi)^{-2d} \sum_j [\{\int_\Omega \gamma(\theta)^2 e^{i\langle j,\theta\rangle}\, d\theta\}\{\int_\Omega \chi(\theta)^{-2} e^{i\langle j,\theta\rangle}\, d\theta\} \\
&\quad + \{\int_\Omega \gamma(\theta)\chi(\theta)^{-1} e^{i\langle j,\theta\rangle}\, d\theta\}^2] \\
&= 2(2\pi)^{-d} \int_\Omega \gamma(\theta)^2 \chi(\theta)^{-2}\, d\theta. \qquad (3.4.52)
\end{aligned}$$

Here J is derived from $\mathbf{E}B^2$ by replacing one sum '$j \in \mathcal{R}$' by '$j \in \mathbb{Z}^d$'. This change to the infinite sum allows us to apply Parseval's identity which then leads to the last integral in (3.4.52).

It remains to estimate (3.4.52). Define $e : \mathbb{Z}^d \to \mathbf{R}$ by

$$e(j) = \begin{cases} -k_\Theta(j) & \text{if } \; j \in \mathcal{N} \\ \kappa \sum_{l\in\mathcal{N}} k_\Theta(l) & \text{if } \; j \in \mathcal{M} \\ 0 & \text{if } \; j \in \widetilde{\mathcal{M}} \cap \widetilde{\mathcal{N}}, \end{cases}$$

where $\kappa = c_{\mathcal{M}}^{-1}$, as defined at the start of the proof. Let η denote the Fourier transform of e. In the notation of (3.4.50) one obtains

$$\begin{aligned}
\alpha_{jl} &= k_\Theta(j-l) + e(j-l) \quad \text{for } j, l \in \mathbb{Z}^d \\
\gamma^2 &= (\chi^{-1}\mathcal{I}_\Theta + \eta)^2,
\end{aligned}$$

and therefore J of (3.4.52) becomes

$$J = 2(2\pi)^{-d}\{\int_\Theta \chi(\theta)^{-4}\, d\theta + 2\int_\Theta \chi(\theta)^{-3}\eta(\theta)\, d\theta + \int_\Omega \chi(\theta)^{-2}\eta(\theta)^2\, d\theta\}. \tag{3.4.53}$$

From Lemma 3.4.2 it follows that

$$\int_\Theta \chi(\theta)^{-4}\, d\theta \asymp \delta^d \max\{1, (\lambda\delta)^{8d}\}. \tag{3.4.54}$$

To bound the other two terms in (3.4.53), observe that

$$\eta(\theta) = \sum_j e(j)e^{i\langle j,\theta\rangle} = \sum_{l\in\mathcal{N}} k_\Theta(l)\{\kappa \sum_{j\in\mathcal{M}} e^{i\langle j,\theta\rangle} - e^{i\langle l,\theta\rangle}\}, \tag{3.4.55}$$

and therefore

$$\begin{aligned}
\int_\Theta \chi(\theta)^{-3}\eta(\theta)\, d\theta &= \sum_{l\in\mathcal{N}} k_\Theta(l) \int_\Theta \chi(\theta)^{-3}\{\kappa \sum_{j\in\mathcal{M}} e^{i\langle j,\theta\rangle} - e^{i\langle l,\theta\rangle}\}\, d\theta \\
&= (2\pi)^{-d}\sum_{l\in\mathcal{N}}\{\int_\Theta \chi(\theta)^{-1}e^{-i\langle l,\theta\rangle}\, d\theta\}[\int_\Theta \chi(\theta)^{-3}\{\kappa\sum_{j\in\mathcal{M}} e^{i\langle j,\theta\rangle} - e^{i\langle l,\theta\rangle}\}\, d\theta] \\
&= O[\delta^{2d}\max\{1,(\lambda\delta)^{8d}\}]
\end{aligned} \tag{3.4.56}$$

follows again by Lemma 3.4.2.

Lastly we estimate $\int_\Omega \chi(\theta)^2\eta(\theta)^2\, d\theta$. Observe that $k_\Theta(l) = k_\Theta(0)\{1+o(1)\}$ for $l \in \mathcal{N}$, and therefore $\eta(\theta)$ given in (3.4.55) becomes

$$\eta(\theta) = k_\Theta(0)\{1+o(1)\}\sum_{l\in\mathcal{N}}\{\kappa\sum_{j\in\mathcal{M}} e^{i\langle j,\theta\rangle} - e^{i\langle l,\theta\rangle}\}.$$

Now,

$$\begin{aligned}
\int_\Omega \chi(\theta)^2\eta(\theta)^2\, d\theta &= k_\Theta(0)^2\{1+o(1)\}\int_\Omega \chi(\theta)^{-2}\sum_{l\in\mathcal{N}}\{\kappa\sum_{j\in\mathcal{M}} e^{i\langle j,\theta\rangle} - e^{i\langle l,\theta\rangle}\}\, d\theta \\
&= O[\lambda^{4d}\delta^{2d}\max\{1,(\lambda\delta)^{4d}\}]
\end{aligned} \tag{3.4.57}$$

by Lemma 3.4.2. Combining (3.4.53), (3.4.54), (3.4.56) and (3.4.57) gives the following bound for J:

$$J = O[\lambda^{4d}\delta^{2d}\max\{1,(\lambda\delta)^{4d}\} + \delta^d\max\{1,(\lambda\delta)^{8d}\}]. \tag{3.4.58}$$

Finally note that the sum '$j \in \mathcal{R}$' in (3.4.51) may be replaced by the infinite sum '$j \in \mathbb{Z}^d$' (the latter was used in the definition of J) without affecting the asymptotics. This can be seen by arguments similar to those given in the proof of Lemma 3.4.3. Making use of the estimate (3.4.58) for J, one now obtains

$$\begin{aligned}
\mathbf{E}T^2 &= O[(n\lambda^4\delta^2)^d\max\{1,(\lambda\delta)^{4d}\} + (n\delta)^d\max\{1,(\lambda\delta)^{8d}\}] \\
&= O[(n\lambda^4\delta^2)^d\max\{1,(\lambda\delta)^{4d}\}] + o(\mathrm{MSSE}^2)
\end{aligned}$$

as desired. □

Proof of Theorem 3.6

For $n > 0$, $\mathcal{R} = \mathcal{R}_n$, N_2, SSE_2^* and CV_2 are given by (3.3.8)–(3.3.10) as

$$\begin{aligned} N_2 &= \sum_{j \in \mathcal{R}} t(j)(\epsilon \star h^{(-1)})(j); \\ \mathrm{SSE}_2^* &= \sum_{j \in \mathcal{R}} \{\hat{t}(j)^2 - 2\hat{t}_2^*(j)\tilde{t}_2(j) + t(j)^2\}; \\ \mathrm{CV}_2 &= \mathrm{SSE}_2^* + N_2. \end{aligned}$$

Here SSE_2^* is the (deconvolved) cross-validation approximation to SSE. Since N_2 contains only elements which do not depend on δ, it suffices to consider $\mathrm{CV}_2 - \mathrm{SSE}$ instead of $\mathrm{SSE}_2^* - \mathrm{SSE}$.

For the remainder of this proof we drop the subscript 2 and write CV, SSE^*, $\hat{t}^*$ and $\tilde{t}$ instead of CV_2, SSE_2^*, $\hat{t}_2^*$ and $\tilde{t}_2$. Now put

$$\begin{aligned} A &= \sum_{j \in \mathcal{R}} \{\hat{t}^*(j) - \hat{t}(j)\} t(j), \\ B &= \sum_{j \in \mathcal{R}} \{\hat{t}^*(j) - t(j)\}(\epsilon \star h^{(-1)})(j). \end{aligned} \tag{3.4.59}$$

Then

$$\mathrm{CV} - \mathrm{SSE} = -2(A + B). \tag{3.4.60}$$

Using this notation, we shall show the following

- $\mathbf{E}A^2 = o(\mathrm{MSSE}^2)$; and
- $\mathbf{E}B^2 = o(\mathrm{MSSE}^2)$ if and only if

$$(\delta\lambda^2 n^{-1/2})^d \{1 + (\lambda\delta)^{2d}\}\{(n\delta)^{d-2a} + \delta^d + (\lambda^4\delta^5)^d\}^{-1} \to 0.$$

We recall that

$$\mathrm{MSSE} \asymp n^d[(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}]$$

was given in Proposition 3.1.

For $i \in \mathbb{Z}^d$ define D_i on points y_j, $j \in \mathbb{Z}^d$ by

$$\mathsf{D}_i y_j = c_{\mathcal{M}}^{-1} \sum_{l \in \mathcal{M}} y_{j+l} - y_{j+i}, \tag{3.4.61}$$

where $\mathcal{M} = \{j \in \mathbb{Z}^d \,:\, \|j\|_\infty = 2\}$ and $c_\mathcal{M} = |\mathcal{M}|$ as in (3.2.24). The estimator $\hat{t}^*$ of (3.2.25) can now be re-written in the following way:

$$\begin{aligned}\hat{t}^*(j) &= \hat{t}(j) + \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i X_j \\ &= t_\Theta(j) + \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i(h\star t)(j) + (\epsilon\star k_\Theta)(j) + \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i\epsilon_j. \quad (3.4.62)\end{aligned}$$

We now estimate $\mathbf{E}A^2$, where A was defined in (3.4.59). Put

$$\begin{aligned}A_1 &= \sum_{j\in\mathcal{R}} t(j) \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i(h\star t)(j), \\ A_2 &= \sum_{j\in\mathcal{R}} t(j) \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i\epsilon_j.\end{aligned}$$

Using (3.4.62), one obtains $A = A_1 + A_2$, and it therefore suffices to estimate A_1^2 and $\mathbf{E}A_2^2$. Consider

$$\begin{aligned}|A_1| &= \sup_{j\in\mathcal{R}} |t(j)| \sup_{i\in\mathcal{N}} |k_\Theta(i)| \sum_{j\in\mathcal{R}}\sum_{i\in\mathcal{N}} |\mathsf{D}_i(h\star t)(j)| \\ &\leq c_1 \int_\Omega \tau(\theta)\, d\theta \int_\Theta \chi(\theta)^{-1}\, d\theta \sum_{j\in\mathcal{R}}\sum_{i\in\mathcal{N}} |\mathsf{D}_i(h\star t)(j)| \qquad (c_1 > 0) \\ &= o[(n\delta)^d \max\{1, (\lambda\delta)^{2d}\}], \qquad (3.4.63)\end{aligned}$$

since the first factor is $O(1)$ by Lemma 3.4.1, the second is $O[\delta^d \max\{1, (\lambda\delta)^{2d}\}]$ by Lemma 3.4.2, and $|\mathsf{D}_i(h\star t)(j)| = o(1)$ by a proof analogous to that given in Lemma 3.4.5.

An estimate for $\mathbf{E}A_2^2$ is obtained in the following argument.

$$\begin{aligned}\mathbf{E}A_2^2 &= \sum_{j_1\in\mathcal{R}}\sum_{j_2\in\mathcal{R}} t(j_1)t(j_2) \sum_{i_1\in\mathcal{N}}\sum_{i_2\in\mathcal{N}} k_\Theta(i_1)k_\Theta(i_2)\mathbf{E}(\mathsf{D}_{i_1}\epsilon_{j_1}\mathsf{D}_{i_2}\epsilon_{j_2}) \\ &\leq c_2 \sup_{j\in\mathcal{R}} |t(j)|^2 \sup_{i\in\mathcal{N}} |k_\Theta(i)|^2 \sum_{j\in\mathcal{R}} \sigma^2 \qquad (c_2 > 0) \\ &= O[(n\delta^2)^d \max\{1, (\lambda\delta)^{4d}\}] \qquad (3.4.64)\end{aligned}$$

by Lemmas 3.4.1 and 3.4.2 and the fact that $\mathbf{E}\epsilon^2 = \sigma^2$. Combining (3.4.63) and (3.4.64), an estimate for $\mathbf{E}A^2$ is given by

$$\begin{aligned}\mathbf{E}A^2 &= o[(n\delta)^{2d}\max\{1, (\lambda\delta)^{4d}\}] + O[(n\delta^2)^d \max\{1, (\lambda\delta)^{4d}\}] \\ &= o(\mathrm{MSSE}^2). \qquad (3.4.65)\end{aligned}$$

We now turn to the second term, B, in CV – SSE (see (3.4.59) and (3.4.60)). If $\mathcal{B}(j)$ denotes the bias of $\hat{t}(j)$, then by (3.4.62), B is given by

$$B = \sum_{j\in\mathcal{R}} \{\mathcal{B}(j) + \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i(h \star t)(j) + (\epsilon \star k_\Theta)(j) + \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i\epsilon_j\}(\epsilon \star h^{(-1)})(j).$$

Put

$$\begin{aligned} B_1 &= \sum_{j\in\mathcal{R}} (\epsilon \star h^{(-1)})(j)\{(\epsilon \star k_\Theta)(j) + \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i\epsilon_j\}; \\ B_2 &= \sum_{j\in\mathcal{R}} \mathcal{B}(j)(\epsilon \star h^{(-1)})(j); \\ B_3 &= \sum_{j\in\mathcal{R}} (\epsilon \star h^{(-1)})(j) \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i(h \star t)(j). \end{aligned} \tag{3.4.66}$$

Then $B = B_1 + B_2 + B_3$, and it suffices to estimate $\mathbf{E}B_i^2$, $i = 1, 2, 3$. By Lemma 3.4.7 (which precedes this proof),

$$\mathbf{E}B_1^2 = O[(n\lambda^4\delta^2)^d \max\{1, (\lambda\delta)^{4d}\}] + o(\mathrm{MSSE}^2). \tag{3.4.67}$$

Consider

$$\mathbf{E}B_2^2 = \sum_{j_1\in\mathcal{R}} \sum_{j_2\in\mathcal{R}} \mathcal{B}(j_1)\mathcal{B}(j_2) \sum_l \mathbf{E}\epsilon_l^2 h^{(-1)}(j_1 - l)h^{(-1)}(j_2 - l),$$

and put

$$I(B_2^2) = \sum_l \{\sum_j \mathcal{B}(j)h^{(-1)}(j - l)\}^2,$$

that is, $I(B_2^2)$ is obtained from $\mathbf{E}B_2^2$ by replacing sums '$j \in \mathcal{R}$' by sums '$j \in \mathbb{Z}^d$'. Now

$$\begin{aligned} I(B_2^2) &= (2\pi)^{-d} \int_{\Omega\setminus\Theta} \tau(\theta)^2 \chi(\theta)^{-2}\, d\theta \\ &= O[n^d(n\delta)^{d-2a} \max\{1, (\lambda\delta)^{4d}\}]. \end{aligned}$$

The first equality follows from Parseval's identity. The next step follows from a slight modification of part 2 of Lemma 3.4.6 (where we have replaced $\chi(\theta)^{-1} - 1$ by $\chi(\theta)^{-1}$). Note that for Lemma 3.4.6 we require that $a > 5d/2$. We may conclude that

$$\mathbf{E}B_2^2 = O[n^d(n\delta)^{d-2a} \max\{1, (\lambda\delta)^{4d}\}] = o(\mathrm{MSSE}^2), \tag{3.4.68}$$

since the asymptotics are unaffected by going from $I(B_2^2)$ to $\mathbf{E}B_2^2$ as can be shown by arguments similar to those given in Lemma 3.4.3.

It remains to estimate

$$\mathbf{E}B_3^2 = \mathbf{E}\{\sum_{j\in\mathcal{R}} (\epsilon \star h^{(-1)})(j) \sum_{i\in\mathcal{N}} k_\Theta(i)\mathsf{D}_i(h \star t)(j)\}^2$$

$$= \sigma^2 \sum_{j_1 \in \mathcal{R}} \sum_{j_2 \in \mathcal{R}} \sum_{l} h^{(-1)}(j_1 - l) h^{(-1)}(j_2 - l) \Upsilon_{j_1} \Upsilon_{j_2},$$

where

$$\Upsilon_j = \sum_{i \in \mathcal{N}} k_\Theta(i) \mathsf{D}_i(h \star t)(j).$$

From Lemma 3.4.5 it follows that $\mathsf{D}_i(h \star t)(j) = o(1)$ by replacing $\sum_{|i|=1}$ by $\sum_{i \in \mathcal{N}}$, a change which does not affect the argument. Furthermore, by Lemma 3.4.2, $\sup_{i \in \mathcal{N}} |k_\Theta(i)| = O[\delta^d \max\{1, (\lambda\delta)^{2d}\}]$, and therefore

$$\Upsilon_j = o[\delta^d \max\{1, (\lambda\delta)^{2d}\}] \quad \text{for } j \in \mathcal{R}. \tag{3.4.69}$$

Returning to $\mathbf{E}B_3^2$, one obtains

$$\begin{aligned} \mathbf{E}B_3^2 &= \sigma^2 \sum_{j_1 \in \mathcal{R}} \sum_{j_2 \in \mathcal{R}} \Upsilon_{j_1} \Upsilon_{j_2} (h^{(-1)} \star h^{(-1)})(j_1 - j_2) \\ &= \sigma^2 \sum_{j_1 \in \mathcal{R}} \sum_{j_2 \in \mathcal{N}_2} \Upsilon_{j_1} \Upsilon_{j_2} (h^{(-1)} \star h^{(-1)})(j_1 - j_2), \end{aligned} \tag{3.4.70}$$

since $h^{(-1)} \star h^{(-1)}$ has support in $\mathcal{N}_2 = \{j_1 + j_2 \ : \ j_1, j_2 \in \mathcal{N}\}$. This set $\mathcal{N}_2$ is finite and independent of n. Furthermore,

$$(h^{(-1)} \star h^{(-1)})(j) = O(\lambda^{4d}) \tag{3.4.71}$$

follows from the fact that $|f \star g| \le \|f\|_\infty \|g\|_1$ for functions $f,\ g \in L^1$, and from Lemma 3.4.2. Combining (3.4.69) -(3.4.71) leads to

$$\mathbf{E}B_3^2 = o[(n\lambda^4\delta^2)^d \max\{1, (\lambda\delta)^{4d}\}] = o(\mathbf{E}B_1^2). \tag{3.4.72}$$

The bounds in (3.4.67), (3.4.68) and (3.4.72) now provide an estimate for $\mathbf{E}B^2$.

$$\mathbf{E}B^2 = O[(n\lambda^4\delta^2)^d \max\{1, (\lambda\delta)^{4d}\}] + o(\text{MSSE}^2). \tag{3.4.73}$$

To complete the proof of the theorem, it remains to establish conditions on λ such that $\mathbf{E}B^2/\text{MSSE}^2 \to 0$. (Recall that $\mathbf{E}A^2 = o(\text{MSSE}^2)$ by (3.4.65).)

By Proposition 3.1, MSSE $\asymp n^d[(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}]$. It thus follows from (3.4.72) that

$$\mathbf{E}B^2/\text{MSSE}^2 \to 0$$

if and only if

$$n^{-d}\lambda^{4d}\delta^{2d} \max\{1, (\lambda\delta)^{4d}\}[(n\delta)^{d-2a} + \delta^d \max\{1, (\lambda\delta)^{4d}\}]^{-2} \to 0.$$

Consider the case where δ is chosen to minimise MSSE. We treat two separate cases: $\lambda\delta \le k$, for some finite constant k, and $\lambda\delta \to \infty$.

Without loss of generality, we may take $k = 1$, and consider $\lambda\delta \le 1$ as $n \to \infty$. In this case, $\delta = c_1\delta_0$ where $\delta_0 = n^{-1+d/2a}$ minimises the order of MSSE and $c_1 > 0$. By Corollary 3.2, $\mathrm{MSSE}(\delta_0) \asymp n^{d^2/2a}$ and thus for $\delta = c_1\delta_0$

$$\mathbf{E}B^2/\mathrm{MSSE}^2 \to 0 \quad \text{if and only if} \quad n^{-d}\lambda^{4d} \to 0,$$

and this is equivalent to $\lambda = o(n^{1/4})$.

Now assume that $\lambda\delta \to \infty$ as $n \to \infty$. Then $\delta_0 = (n^{d-2a}\lambda^{-4d})^{1/(4d+2a)}$ minimises the order of MSSE, and $\mathrm{MSSE}(\delta_0) \asymp n^d(n^{-5}\lambda^4)^{d(2a-d)/(4d+2a)}$ by Corollary 3.2. This implies that for $\delta = c_2\delta_0$ $(c_2 > 0)$,

$$\begin{aligned}\mathbf{E}B^2/\mathrm{MSSE}^2 &\asymp \{n^{-d}\lambda^{8d}\delta^{6d}(n\delta)^{4a-2d}\}\\ &= n^{3a-4d}\lambda^{8d}.\end{aligned}$$

Now

$$n^{3a-4d}\lambda^{8d} \to 0 \quad \text{if and only if} \quad \lambda = o(n^{(4d-3a)/8d}).$$

But $\lambda\delta > k$ together with $\delta = c_2\delta_0$ imply that

$$n^{1-d/2a} = O(\lambda) = o(n^{(4d-3a)/8d}).$$

From this last relationship one deduces that $3a^2 - 4d^2 + 4ad \le 0$, which contradicts the assumption that $a > d$. Therefore the case $\lambda\delta \to \infty$ is excluded.

This completes the proof of Theorem 3.6. □

3.4.6 Proof of Proposition 3.7

Proposition 3.7 *Assume that t, h and ϵ satisfy* **A1–A4**. *Assume further that $a > 3d$ and $\mathbf{E}\epsilon^4 < \infty$. If $\gamma > 2 + 2a - d/2a$, then*

$$\sup_{0\le i<\kappa}\sup_{0<r\le 1}\left|\frac{CV_1(n,\rho_i) - SSE(n,\rho_i)}{MSSE(n,\rho_i)} - \frac{CV_1(n,\phi_{i,r}) - SSE(n,\phi_{i,r})}{MSSE(n,\phi_{i,r})}\right| = o(n^{-1}) \quad a.s.$$

We first prove two lemmas. Apart from Lemmas 3.4.8 and 3.4.9 below, the proof uses Lemmas 3.4.1–3.4.3 and 3.4.5–3.4.6 as well as results from the proof of Theorem 3.5. The notation we use is that established in the proof of Theorem 3.5, and some of our estimates are just those obtained in the proof of the mean square results (Theorem 3.5). One of the main differences to the proof of Theorem 3.5, however, is that we now regard the estimator $\hat{t}$ (see (3.3.2)) explicitly as a function of the smoothing parameter.

To do this, we shall use the following notation for $x \in I$:

$$\begin{aligned}
\hat{t}_x(j) &= (2\pi)^{-d} \int_{\Theta_x} \tau(\theta) e^{-i\langle j,\theta\rangle}\, d\theta + (k_x \star \epsilon)(j); \\
k_x(j) &= (2\pi)^{-d} \int_{\Theta_x} \chi(\theta)^{-1} e^{-i\langle j,\theta\rangle}\, d\theta; \\
\mathcal{B}_x(j) &= -(2\pi)^{-d} \int_{\Omega\setminus\Theta_x} \tau(\theta) e^{-i\langle j,\theta\rangle}\, d\theta; \\
\mathcal{V}_x(j) &= \mathbf{E}\{(k_x \star \epsilon)(j)^2\}, \qquad (3.4.74)
\end{aligned}$$

where $\Theta_x = \{\theta \in \Omega \ : \ |\theta_i| \le x \ \ \forall i = 1, \ldots, d\}$. If no specific smoothing parameter is considered, we shall omit the subscript and revert to the previous notation $\hat{t}$, k_Θ, $\mathcal{B}$ and $\mathcal{V}$.

Lemma 3.4.8 *For $\gamma > 1 - d/2a$, $\kappa \le k_3 n^{\gamma-1+d/2a}$ with k_3 as in (3.3.16), put*

$$\rho_i = k_1\delta_0 + in^{-\gamma}, \qquad \phi_{i,r} = \rho_i + rn^{-\gamma} \quad \text{for } 0 \le i < \kappa \ \text{ and } 0 < r \le 1.$$

If $\alpha \in \mathbf{R}$, then

1. $|\rho_i - \phi_{i,r}|^\alpha = r^\alpha n^{-\alpha\gamma}$;
2. $|\rho_i^\alpha - \phi_{i,r}^\alpha| \asymp n^{-\gamma+(1-\alpha)(1-d/2a)}$.

Proof of Lemma 3.4.8

Choose an integer i, $0 \le i < \kappa$. Put $y = \rho_i$, $x = \phi_{i,r}$. Now, $|x - y| = rn^{-\gamma}$, and thus for $\alpha \in \mathbf{R}$,

$$|x - y|^\alpha = r^\alpha n^{-\alpha\gamma}.$$

To show part 2, note that $yn^\gamma = k_1 n^{\gamma-1+d/2a} + i$ implies that

$$k_1 n^{\gamma-1+d/2a} \le yn^\gamma \le (k_1 + k_3) n^{\gamma-1+d/2a},$$

since $i < \kappa \le k_3 n^{\gamma-1+d/2a}$. It follows that

$$\begin{aligned}
yn^\gamma &\asymp n^{\gamma-1+d/2a} \quad \text{and} \\
(yn^\gamma)^{-1} &\asymp n^{-\gamma+1-d/2a}.
\end{aligned}$$

Thus for $\alpha \in \mathbf{R}$,

$$\begin{aligned}
|y^\alpha - x^\alpha| &= y^\alpha |1 - (1 + ry^{-1}n^{-\gamma})^\alpha| \\
&\asymp y^\alpha |1 - (1 + rn^{-\gamma+1-d/2a})^\alpha| \\
&= y^\alpha |\alpha r n^{-\gamma+1-d/2a} + O\{n^{2(-\gamma+1-d/2a)}\}| \\
&= |\alpha| r y^\alpha n^{-\gamma+1-d/2a}(1 + o(1)),
\end{aligned}$$

since $\gamma > 1 - d/2a$. But $y \asymp \delta_0 = n^{-1+d/2a}$, and therefore

$$y^\alpha n^{-\gamma+1-d/2a} \asymp n^{(-1+d/2a)\alpha} n^{-\gamma+1-d/2a} = n^{-\gamma+(1-\alpha)(1-d/2a)},$$

from which the desired result follows since $r > 0$. □

Lemma 3.4.9 *For $\gamma > 1 - d/2a$, $\kappa \le k_3 n^{\gamma-1+d/2a}$ with k_3 as in (3.3.16). Put*

$$\rho_i = k_1\delta_0 + in^{-\gamma}, \qquad \phi_{i,r} = \rho_i + rn^{-\gamma} \quad \textit{for } 0 \le i < \kappa \textit{ and } 0 < r \le 1.$$

Then, as $n \to \infty$,

1. $MSSE(\phi_{i,r}) \asymp (n\delta_0)^d = n^{d^2/2a}$;
2. $|MSSE(\phi_{i,r}) - MSSE(\rho_i)| \asymp n^{d-\gamma}\{n^{(1-d+2a)(1-d/2a)} + n^{(1-d)\gamma}\}$.

Proof of Lemma 3.4.9

For $0 \le i < \kappa$, $0 < r \le 1$, put $y = \rho_i$ and $x = \phi_{i,r}$. Since $x \asymp \delta_0$, it follows from Proposition 3.1 that

$$\text{MSSE}(x) \asymp n^d\{(nx)^{d-2a} + x^d\} \asymp n^d\{(n\delta_0)^{d-2a} + \delta_0^d\} = n^{d^2/2a}.$$

To show part 2, observe that

$$\text{MSSE}(x) - \text{MSSE}(y) = \sum_{j\in\mathcal{R}}\{\mathcal{B}_x(j)^2 - \mathcal{B}_y(j)^2\} + \sum_{j\in\mathcal{R}}\{\mathcal{V}_x(j) - \mathcal{V}_y(j)\},$$

where $\mathcal{B}_\delta$ and $\mathcal{V}_\delta$ are given by (3.4.74). Consider

$$\begin{aligned} |\sum_{j\in\mathcal{R}}\{\mathcal{V}_x(j) - \mathcal{V}_y(j)\}| &= |n^d\sigma^2\sum_l\{k_x(l)^2 - k_y(l)^2\}| \\ &= |n^d\sigma^2\int_{\Theta_x\setminus\Theta_y}\chi(\theta)^{-2}\,d\theta| \\ &\asymp n^d|x-y|^d. \end{aligned}$$

The second last equality follows from Parseval's identity and the last $\asymp$-relationship follows from Lemma 3.4.2. For the bias term observe that

$$|\sum_{j\in\mathcal{R}}\{\mathcal{B}_x(j)^2 - \mathcal{B}_y(j)^2\}| \asymp |\int_{\Theta_x\setminus\Theta_y}\tau(\theta)^2\,d\theta| \asymp n^d|x^{d-2a} - y^{d-2a}|$$

follows from Lemmas 3.4.3 and 3.4.1. Therefore one has

$$\begin{aligned} |\text{MSSE}(x) - \text{MSSE}(y)| &\asymp n^d\{|x^{d-2a} - y^{d-2a}| + |x-y|^d\} \\ &\asymp n^d\{n^{-\gamma+(1-d+2a)(1-d/2a)} + n^{-d\gamma}\} \end{aligned}$$

$$= n^{d-\gamma}\{n^{(1-d+2a)(1-d/2a)} + n^{(1-d)\gamma}\}$$

as required. □

Proof of Proposition 3.7

Fix $n \in \mathbb{N}$. For notational convenience, we put $\mathcal{C} = \mathrm{CV}_1 - \mathrm{SSE}$ and $\mathcal{M} = \mathrm{MSSE}$ here. Let x, y denote points in $I = [k_1\delta_0,\ k_2\delta_0]$. To prove the proposition, we use the following inequality:

$$\left|\frac{\mathcal{C}(x)}{\mathcal{M}(x)} - \frac{\mathcal{C}(y)}{\mathcal{M}(y)}\right| \leq |\mathcal{C}(x) - \mathcal{C}(y)||\mathcal{M}(x)|^{-1} + |\mathcal{C}(y)||\mathcal{M}(x)\mathcal{M}(y)|^{-1}|\mathcal{M}(x) - \mathcal{M}(y)|. \tag{3.4.75}$$

The first part of this proof is concerned with estimates for $\mathcal{C}(x) - \mathcal{C}(y)$ and $\mathcal{C}(y)$ for points $x, y \in I$. Assume that $x > y$. To obtain estimates for $\mathcal{C}(x) - \mathcal{C}(y)$ and $\mathcal{C}(y)$ put

$$\mathcal{C} = -2(U + V + W), \tag{3.4.76}$$

where

$$\begin{aligned}
U &= \sum_{j\in\mathcal{R}} \{\hat{t}^*(j) - \hat{t}(j)\}(h\star t)(j) \\
&= \sum_{j\in\mathcal{R}} \{(h\star t)(j)\mathsf{D}(h\star t)(j) + (h\star t)(j)\mathsf{D}\epsilon_j\}k_\Theta(0); \\
V &= \sum_{j\in\mathcal{R}} \{\hat{t}^*(j) - t(j)\}\epsilon_j \\
&= \sum_{j\in\mathcal{R}} [\{\mathcal{B}(j) + k_\Theta(0)\mathsf{D}(h\star t)(j)\}\epsilon_j + \{(k_\Theta\star\epsilon)(j) + k_\Theta(0)\mathsf{D}\epsilon_j\}\epsilon_j]; \\
W &= \sum_{j\in\mathcal{R}} \{\hat{t}(j) - t(j) - (\epsilon\star h^{(-1)})(j)\}\{(h\star t)(j) - t(j)\} \\
&= \sum_{j\in\mathcal{R}} \{\mathcal{B}(j) - (\epsilon\star h^{(-1)})(j)\}\{(h\star t)(j) - t(j)\}
\end{aligned} \tag{3.4.77}$$

as in (3.4.28), (3.4.33) and (3.4.41) in the proof of Theorem 3.5, and where

$$\mathsf{D}z_j = (2d)^{-1}\sum_{|l|=1} z_{j+l} - z_j$$

as in (3.4.27). Now

$$|k_x(0) - k_y(0)| = (2\pi)^{-d}\left|\int_{\Theta_x\setminus\Theta_y} \chi(\theta)^{-1}\, d\theta\right| \asymp |x - y|^d \tag{3.4.78}$$

follows from the proof of Lemma 3.4.2, since $\lambda x < 1$ and

$$\Theta_x \setminus \Theta_y = \{\theta \in \Omega\ :\ y < |\theta_i| \leq x\ \forall i = 1, \ldots, d\}.$$

As in (3.4.28), let U_1 denote the deterministic part and U_2 the random part of U. Combining (3.4.78) and (3.4.30), one obtains the following estimates:

$$\begin{aligned} U_1(x) - U_1(y) &= o(n^d|x-y|^d); \\ U_1(y) &= o(n^d y^d). \end{aligned} \tag{3.4.79}$$

To estimate U_2, consider

$$u_n = \sum_{j\in\mathcal{R}} (h \star t)(j)\mathsf{D}\epsilon_j = \sum_{j\in\mathcal{R}^+} a_j\epsilon_j,$$

where $\mathcal{R}^+ = \{k \in \mathbb{Z}^d \ : \ k = j+l, \ j \in \mathcal{R}, \ |l| = 1\}$ and the a_j denote finite sums of the form $\sum_{|l|\le 1} \alpha_{j+l}(h\star t)(j+l)$ for finite α_{j+l}. In fact, for most $j \in \mathcal{R}$, $a_j = \mathsf{D}(h\star t)(j)$. The ϵ_j are uncorrelated, and $\sup_{j\in\mathcal{R}^+} \mathbf{E}|a_j\epsilon_j|^2 = O(1)$ by (3.4.29). It therefore follows by the strong law of large numbers (see Theorem 5.1.2 of Chung (1974)) that

$$n^{-d}u_n \to 0 \quad a.s. \tag{3.4.80}$$

Together with (3.4.78), (3.4.80) yields that

$$\begin{aligned} U_2(x) - U_2(y) &= o(n^d|x-y|^d) \quad a.s.; \\ U_2(y) &= o(n^d y^d) \quad a.s. \end{aligned} \tag{3.4.81}$$

To obtain an estimate for V, first consider

$$\begin{aligned} V_{11}(x) - V_{11}(y) &\equiv \sum_{j\in\mathcal{R}} \{\mathcal{B}_x(j) - \mathcal{B}_y(j)\}\epsilon_j \\ &= (2\pi)^{-d} \sum_{j\in\mathcal{R}} \epsilon_j \int_{\Theta_x\setminus\Theta_y} \tau(\theta)e^{-i\langle j,\theta\rangle}\, d\theta, \end{aligned}$$

where V_{11} is the first part of V_1 (see (3.4.34)). A straightforward adaptation of Lemma 3.4.1 shows that

$$\left|\int_{\Theta_x\setminus\Theta_y} \tau(\theta)e^{-i\langle j,\theta\rangle}\, d\theta\right| \asymp n^{d-a}|x^{d-a} - y^{d-a}|.$$

Now put

$$Z_j = \epsilon_j\{n^{d-a}|x^{d-a} - y^{d-a}|\}^{-1} \int_{\Theta_x\setminus\Theta_y} \tau(\theta)e^{-i\langle j,\theta\rangle}\, d\theta,$$

and observe that the Z_j are independent and $\mathbf{E}Z_j^2 \le \sigma^2$. By Theorem 5.1.2 of Chung (1974) it therefore follows that

$$\sum_{j\in\mathcal{R}} Z_j = o(n^d) \quad a.s.,$$

and thus one may conclude that

$$\begin{aligned} V_{11}(x) - V_{11}(y) &= o(n^{2d-a}|x^{d-a} - y^{d-a}|) \quad a.s.; \\ V_{11}(y) &= o(n^{2d-a}y^{d-a}) \quad a.s. \end{aligned} \tag{3.4.82}$$

Let $V_{12} = V_1 - V_{11}$, then

$$V_{12} = \sum_{j\in\mathcal{R}} k_\Theta(0)\mathsf{D}(h \star t)(j)\epsilon_j = o(n^d\delta^d) \quad a.s. \tag{3.4.83}$$

Here we have used the fact that V_{12} is of the same form as U_2, the random part of U, as given in (3.4.77). The same estimate therefore applies to V_{12}.

Next let V_{21} denote the first summand in (3.4.37). Then

$$V_{21}(x) - V_{21}(y) = d^{-1}\{k_x(0) - k_y(0)\}\sum_{l\in\mathcal{P}}\sum_{j\in\mathcal{R}} \epsilon_j\epsilon_{j+l}, \tag{3.4.84}$$

where $\mathcal{P} = \{l \in \mathbb{Z}^d \ : \ |l| = 1, \ l_i \le 0 \ \forall i = 1,\ldots,d\}$. The random variables $\epsilon_j\epsilon_{j+l}$ are independent, and it therefore follows from Theorem 5.1.2 of Chung (1974) and (3.4.78) that

$$\begin{aligned} V_{21}(x) - V_{21}(y) &= o(n^d|x-y|^d) \quad a.s.; \\ V_{21}(y) &= o(n^d y^d) \quad a.s. \end{aligned} \tag{3.4.85}$$

Put $V_{22} = V_2 - V_{21}$ with V_2 as defined in (3.4.37). For x and y one has

$$V_{22}(x) - V_{22}(y) = \sum_{j\in\mathcal{R}}\sum_{l\neq j} \epsilon_j\epsilon_l\beta_{jl} = \sum_{j\in\mathcal{R}}(2\Phi_j + c_j\Psi_j) \tag{3.4.86}$$

where

$$\begin{aligned} \beta_{jl} &= k_x(j-l) - k_y(j-l), \\ \Phi_j &= \sum_{l\prec j} \epsilon_j\epsilon_l\beta_{jl}, \qquad \Psi_j = \sum_{l\succ j} \epsilon_j\epsilon_l\beta_{jl}, \end{aligned}$$

$c_j = 1$ if $l-j \in \mathcal{R}$ and $c_j = 2$ if $l-j \notin \mathcal{R}$. Here $\prec$ denotes the ordering in $\mathbb{Z}^d$, introduced in the Appendix. Clearly Φ_j, Φ_k and Ψ_j, Ψ_k are uncorrelated, since

$$\begin{aligned} \mathbf{E}(\Phi_j\Phi_k) &= \sum_{l\prec j}\sum_{m\prec k} \mathbf{E}(\epsilon_j\epsilon_l\epsilon_m\epsilon_k)\beta_{jl}\beta_{km} \\ &= \begin{cases} \sigma^4\sum_{l\prec j}\beta_{jl}^2 & \text{if } j = k \\ 0 & \text{if } j \neq k. \end{cases} \end{aligned} \tag{3.4.87}$$

Furthermore,

$$\sum_{l\prec j}\beta_{jl}^2 \le \sum_l \{k_x(l)-k_y(l)\}^2 = (2\pi)^{-d}\int_{\Theta_x\setminus\Theta_y}\chi(\theta)^{-2}\,d\theta \asymp |x-y|^d$$

follows from Parseval's identity and Lemma 3.4.2, and thus

$$\mathbf{E}\Phi_j^2 \le c_1|x-y|^d \quad \text{for } c_1>0.$$

Similarly one shows that $\mathbf{E}(\Psi_j\Psi_k)=0$ if $j\neq k$, and $\mathbf{E}\Psi_j^2\le c_2|x-y|^d$. Next put

$$\tilde{\Phi}_j = |x-y|^{-d/2}\Phi_j, \qquad \tilde{\Psi}_j = |x-y|^{-d/2}\Psi_j,$$

then

$$\sum_{j\in\mathcal{R}}(\tilde{\Phi}_j+\tilde{\Psi}_j) = o(n^d) \quad a.s.,$$

again by Theorem 5.1.2 of Chung (1974). From the last equations it now follows that

$$\begin{aligned} V_{22}(x)-V_{22}(y) &= o(n^d|x-y|^{d/2}) \quad a.s.; \\ V_{22}(y) &= o(n^d y^{d/2}) \quad a.s. \end{aligned} \tag{3.4.88}$$

Combining (3.4.79), (3.4.81), (3.4.82), (3.4.85) and (3.4.88), one obtains

$$\begin{aligned} (U+V)(x)-(U+V)(y) &= o(n^{2d-a}|x^{d-a}-y^{d-a}|+n^d|x-y|^{d/2}) \quad a.s.; \\ (U+V)(y) &= o(n^{2d-a}y^{d-a}+n^d y^{d/2}) \quad a.s., \end{aligned} \tag{3.4.89}$$

since $|x-y|\to 0$, $y\to 0$, and therefore $|x-y|=o(|x-y|^{1/2})$ and $y=o(y^{1/2})$.

Now let $0\le i<\kappa$ and $0<r\le 1$. Take $y=\rho_i$ and $x=\phi_{i,r}$ (see (3.5.4) for a definition of ρ_i and $\phi_{i,r}$). By Lemma 3.4.8, $|x-y|^\alpha = r^\alpha n^{-\alpha\gamma}$ and $|x^\alpha - y^\alpha| \asymp n^{-\gamma+(1-\alpha)(1-d/2a)}$. Substitution of these expressions into (3.4.89) yields

$$\begin{aligned} (U+V)(\phi_{i,r})-(U+V)(\rho_i) &= o\{n^{2d-a-\gamma+(1-d+a)(1-d/2a)}+n^{d-d\gamma/2}\} \\ &= o\{n^{d/2-\gamma+1+(d-1)d/2a}+n^{d-d\gamma/2}\} \quad a.s. \end{aligned} \tag{3.4.90}$$

For the second equation in (3.4.89) observe that $y\in I$ implies $y\asymp\delta_0=n^{-1+d/2a}$, and therefore

$$(U+V)(\rho_i) = o\{n^{(1+d/a)d/2}\} \quad a.s. \tag{3.4.91}$$

To obtain an estimate for W, let W_1 denote the deterministic part of W as in (3.4.41):

$$W_1 = \sum_{j\in\mathcal{R}} \mathcal{B}(j)\{(h\star t)(j)-t(j)\}.$$

For $x \in I$, W_1 depends on x only through the bias term $\mathcal{B}$, and thus

$$W_1(x) - W_1(y) = \sum_{j \in \mathcal{R}} \{\mathcal{B}_x(j) - \mathcal{B}_y(j)\}\{(h \star t)(j) - t(j)\}.$$

Since

$$\mathcal{B}_x(j) - \mathcal{B}_y(j) = (2\pi)^{-d} \int_{\Theta_x \setminus \Theta_y} \tau(\theta) e^{-i\langle j,\theta \rangle} \, d\theta,$$

an adaptation of Lemma 3.4.6 to the volume of integration $\Theta_x \setminus \Theta_y$ instead of $\Omega \setminus \Theta$ leads to

$$\begin{aligned} |W_1(x) - W_1(y)| &\asymp n^{2(d-a)} \lambda^2 |x^{d+2-2a} - y^{d+2-2a}|; \\ W_1(y) &\asymp n^{2(d-a)} \lambda^2 y^{d+2-2a} \end{aligned} \tag{3.4.92}$$

(see also (3.4.46)).

As in (3.4.89), we now let $0 \le i < \kappa$ and $0 < r \le 1$, and put $y = \rho_i$, $x = \phi_{i,r}$. The reader may recall from **A4** that $\lambda = O(n^{1-\eta-d/2a})$. Applying the results of Lemma 3.4.8, we get

$$\begin{aligned} |W_1(\phi_{i,r}) - W_1(\rho_i)| &= O\{n^{2(d-a)} n^{2-2\eta-d/a} n^{-\gamma+(-1-d+2a)(1-d/2a)}\} \\ &= O\{n^{1-\gamma-2\eta+(d-1)d/2a}\} \quad a.s.; \end{aligned} \tag{3.4.93}$$

$$W_1(\rho_i) = O\{n^d (n\rho_i)^{d-2a} (\lambda \rho_i)^2\} = O(n^{-2\eta+d^2/2a}) \quad a.s. \tag{3.4.94}$$

Lastly, let W_2 denote the random part of W (see (3.4.41)). Using the notation of (3.4.40), one obtains for $x > y \in I$ that

$$\begin{aligned} W_2(x) - W_2(y) &= \sum_{j \in \mathcal{R}} \{(\epsilon \star r_x)(j) - (\epsilon \star r_y)(j)\} s(j) \\ &= \sum_{j \in \mathcal{R}} \sum_{l} \epsilon_l s(j) \{r_x(j-l) - r_y(j-l)\} \\ &= \sum_{l} \epsilon_l \xi_{l,n}, \end{aligned} \tag{3.4.95}$$

where $\xi_{l,n} = \sum_{j \in \mathcal{R}} s(j)\{r_x(j-l) - r_y(j-l)\}$. Replacing $\xi_{l,n}$ by its limit $\xi_l = \sum_j s(j)\{r_x(j-l) - r_y(j-l)\}$ in (3.4.93), one obtains from (3.4.43) that

$$\mathbf{E}\{\sum_l \epsilon_l \xi_l\}^2 = \sigma^2 \sum_l \xi_l^2. \tag{3.4.96}$$

By applying Parseval's identity and by adapting Lemma 3.4.6 to integration over $\Theta_x \setminus \Theta_y$ instead of $\Omega \setminus \Theta$, one then gets

$$\mathbf{E}\{\sum_l \epsilon_l \xi_l\}^2 \asymp n^{2(d-a)} \lambda^4 |x^{d+4-2a} - y^{d+4-2a}|. \tag{3.4.97}$$

Take $m \in \mathbb{N}$. Equations (3.4.96) and (3.4.97) imply that

$$\begin{aligned} \mathbf{E}\{m^{-d} \sum_{l \in \mathcal{R}_m} \epsilon_l \xi_l\}^2 &\leq \mathbf{E}\{m^{-d} \sum_l \epsilon_l \xi_l\}^2 \\ &\leq c_1 m^{-2d} n^{2(d-a)} \lambda^4 |x^{d+4-2a} - y^{d+4-2a}| \end{aligned}$$

for $c_1 > 0$. In particular, if $m \geq n,\ y = \rho_i$, and $x = \phi_{i,r}$ for $0 \leq i < \kappa,\ 0 < r \leq 1$, then an application of Lemma 3.4.8 leads to

$$\mathbf{E}\{m^{-d} \sum_{l \in \mathcal{R}_m} \epsilon_l \xi_l\}^2 \leq c_2 (n/m)^{2d} n^{1-\gamma-2d-4\eta+(d-1)d/2a} \to 0, \tag{3.4.98}$$

since $\gamma > 2 + 2a - d/2a$ and $\eta > 0$ by assumption. But the ϵ_l are independent, and we may therefore conclude that

$$m^{-d} \sum_{l \in \mathcal{R}_m} \epsilon_l \xi_l \to 0 \quad a.s. \tag{3.4.99}$$

Combining (3.4.98) and (3.4.99) now yields that

$$\sum_{l \in \mathcal{R}_m} \epsilon_l \xi_l = o\{n^{(1-\gamma)/2-2\eta+(d-1)d/4a}\} \quad a.s. \tag{3.4.100}$$

The term $W_2(y)$ may be treated similarly: put $\zeta_{l,n} = \sum_{j \in \mathcal{R}} s(j) r_y(j-l)$ and $\zeta_l = \lim_{n\to\infty} \zeta_{l,n}$. In this notation one may show that

$$\mathbf{E}\{\sum_l \epsilon_l \zeta_l\}^2 \asymp n^{2(d-a)} \lambda^4 y^{d+4-2a}.$$

For $m \geq n,\ y = \rho_i$, as above, it follows that

$$\begin{aligned} \mathbf{E}\{m^{-d} \sum_{l \in \mathcal{R}_m} \epsilon_l \zeta_l\}^2 &\leq c_3 (n/m)^{2d} (ny)^{-2a} (\lambda y)^4 y^d \\ &\leq c_4 (n/m)^{2d} n^{-2d+d^2/2a} n^{-4\eta}, \end{aligned}$$

where we have used that $\rho_i \asymp \delta_0 = n^{-1+d/2a}$, and therefore

$$\sum_{l \in \mathcal{R}_m} \epsilon_l \zeta_l = o(n^{-2\eta+d^2/4a}) \quad a.s. \tag{3.4.101}$$

Replacing ξ_l by $\xi_{l,n}$ and ζ_l by $\zeta_{l,n}$ and the sums over '$l \in \mathcal{R}_m$' by '$l \in \mathbb{Z}^d$' does not affect the asymptotic results, as can be seen by arguments similar to those given in the proof of Lemma 3.4.3. We may therefore conclude that

$$\begin{aligned} W_2(\rho_i) - W_2(\phi_{i,r}) &= o\{n^{-2\eta+(1-\gamma)/2+(d-1)d/4a}\} \quad a.s.; \\ W_2(\rho_i) &= o(n^{-2\eta+d^2/4a}) \quad a.s. \end{aligned} \tag{3.4.102}$$

Combining the bounds obtained for W_1 and W_2 (see (3.4.93), (3.4.94) and (3.4.102)), W is bounded by

$$\begin{aligned} W(\rho_i)-W(\phi_{i,r}) &= O\{n^{1-\gamma-2\eta+(d-1)d/2a}\}+o\{n^{-2\eta+(1-\gamma)/2+(d-1)d/4a}\} \quad a.s.; \\ W(\rho_i) &= O(n^{-2\eta+d^2/2a}) \quad a.s. \end{aligned} \tag{3.4.103}$$

Having calculated bounds for U, V and W, we now return to (3.4.75) and (3.4.76). Combining (3.4.90) and (3.4.103) leads to the following bound for $\mathcal{C}(\rho_i)-\mathcal{C}(\phi_{i,r})$:

$$\begin{aligned} \mathcal{C}(\rho_i)-\mathcal{C}(\phi_{i,r}) &= O\{n^{1-\gamma-2\eta+(d-1)d/2a}\}+o\{n^{(1-\gamma)/2-2\eta+(d-1)d/4a}\} \\ &\quad +o\{n^{1-\gamma+d/2+(d-1)d/2a}+n^{d-d\gamma/2}\} \quad a.s. \end{aligned} \tag{3.4.104}$$

Similarly, by means of (3.4.91) and (3.4.103) one obtains:

$$\mathcal{C}(\rho_i)=o\{n^{(1+d/a)d/2}\}+O(n^{-2\eta+d^2/2a}) \quad a.s. \tag{3.4.105}$$

We now turn to $\mathcal{M}$ = MSSE. Observe that $\mathcal{M}(\phi_{i,r})^{-1} \asymp n^{-d^2/2a}$ and $|\mathcal{M}(\phi_{i,r})-\mathcal{M}(\rho_i)| \asymp n^{d-\gamma}\{n^{(1-d+2a)(1-d/2a)}+n^{(1-d)\gamma}\}$ by Lemma 3.4.9. These estimates together with the bounds given in (3.4.104) and (3.4.105) enable us to show that

$$\begin{aligned} |\mathcal{C}(\rho_i)-\mathcal{C}(\phi_{i,r})|\mathcal{M}(\phi_{i,r})^{-1} &\le c_4 n^{-d^2/2a}\{n^{1-\gamma+d/2+(d-1)d/2a}+n^{d-d\gamma/2} \\ &\quad +n^{1-\gamma-2\eta+(d-1)d/2a}+n^{(1-\gamma)/2-2\eta+(d-1)d/4a}\} \\ &= o(n^{-1}) \quad a.s. \text{ for } c_4>0, \end{aligned} \tag{3.4.106}$$

since $\gamma > 2+2a-d/2a$. Similarly,

$$\begin{aligned} &|\mathcal{C}(\rho_i)||\mathcal{M}(\phi_{i,r})\mathcal{M}(\rho_i)|^{-1}|\mathcal{M}(\phi_{i,r})-\mathcal{M}(\rho_i)| \\ &\le c_5 n^{d-\gamma-d^2/a}\{n^{(1+d/a)d/2}+n^{-2\eta+d^2/2a}\}\{n^{(1-d+2a)(1-d/2a)}+n^{(1-d)\gamma}\} \\ &= o(n^{-1}) \quad a.s. \text{ for } c_5>0. \end{aligned} \tag{3.4.107}$$

Combining (3.4.75), (3.4.106) and (3.4.107), one obtains

$$\sup_{0\le i<\kappa}\sup_{0<r\le 1}\left|\frac{\mathcal{C}(\phi_{i,r})}{\mathcal{M}(\phi_{i,r})}-\frac{\mathcal{C}(\rho_i)}{\mathcal{M}(\rho_i)}\right| = o(n^{-1}) \quad a.s. \text{ as } n\to\infty$$

as required. □

3.4.7 Proof of Theorem 3.8

Theorem 3.8 *Assume that t, h and ϵ satisfy* **A1–A5** *with $\Delta \ge d/4a$. Assume further that $a > 3d$ and $\gamma > 2+2a-d/2a$. If*

1. $b \geq (\gamma + \omega + d/2a) \max\{2a/d^2,\ (4\eta)^{-1}\}$ *for* $\omega > 0$, *and*

2. $\mathbf{E}|\epsilon|^r < \infty$ *for* $1 \leq r \leq 2b$,

then

$$\sup_{\delta \in I} \left| \frac{CV(n,\delta) - SSE(n,\delta)}{MSSE(n,\delta)} \right| \to 0 \quad a.s. \quad as\ n \to \infty.$$

We precede the proof of the theorem with Lemmas 3.4.10 and 3.4.11. The proof also requires Lemmas 3.4.1–3.4.6 given in Subsection 3.4.1, as well as results obtained in the proof of Theorem 3.5.

Lemma 3.4.10 *For* $n > 0$, $\rho \in I = [k_1\delta_0,\ k_2\delta_0]$ *put* $V = \sum_{j\in\mathcal{R}} \sum_{i\neq j} k_{\Theta_\rho}(i-j)\epsilon_i\epsilon_j$. *If* $b \in \mathbb{N}$ *and* $\mathbf{E}|\epsilon|^{2b} < \infty$, *then, as* $n \to \infty$,

$$\sum_{j\in\mathcal{R}} \mathbf{E}\{ \sum_{i\in\mathcal{R};i\neq j} k_{\Theta_\rho}(i-j)\epsilon_i\epsilon_j\}^{2b} = O(n^d \rho^{bd}).$$

Proof of Lemma 3.4.10

For $i, j \in \mathbb{Z}^d$, $\rho \in I$, put $\Theta = \Theta_\rho$ and $\alpha_{ij} = k_\Theta(i-j)$. Consider

$$\begin{aligned} \mathbf{E}|\sum_{i\in\mathcal{R};i\neq j} \alpha_{ij}\epsilon_i\epsilon_j|^{2b} &\leq \sum_{i_1\in\mathcal{R};i_1\neq j} \cdots \sum_{i_{2b}\in\mathcal{R};i_{2b}\neq j} \mathbf{E}|\epsilon_{i_1}\ldots\epsilon_{i_{2b}}|\mathbf{E}\epsilon_j^{2b} \prod_{i=1}^{2b} |\alpha_{ij}| \\ &\leq c_1 \sum_{p=1}^{b} \prod_{m=1}^{p} \sum_{i_m} |\alpha_{i_m j}|^{s_m}, \end{aligned} \tag{3.4.108}$$

where $2 \leq s_m \leq 2b$, $\sum_{m=1}^{p} s_m = 2b$ and the moments $\mathbf{E}|\epsilon_k|^{s_m}\mathbf{E}\epsilon_j^{2b}$ are finite and have been absorbed into the constant $c_1 > 0$. Now, $s_m = 2r$ or $s_m = 2r+1$ for $1 \leq r \leq b$, and therefore

$$\sum_i |\alpha_{ij}|^{s_m} \begin{cases} = \sum_i |\alpha_{ij}|^{2r} & \text{if } s_m = 2r \\ \leq \{\sum_i |\alpha_{ij}|^{4r}\}^{1/2}\{\sum_i |\alpha_{ij}|^2\}^{1/2} & \text{if } s_m = 2r+1, \end{cases} \tag{3.4.109}$$

where the inequality follows from Hölder's inequality. To obtain a bound for (3.4.108), it therefore suffices to estimate sums of the form $\sum_i |\alpha_{ij}|^{2r}$ for $r = 1, \ldots, 2b$. We claim that

$$\sum_i |\alpha_{ij}|^{2r} = O\{\rho^{d(2r-1)}\}. \tag{3.4.110}$$

To see (3.4.110), first note that $\sum_i |\alpha_{ij}| = \sum_i |\alpha_i|$ follows form the definition of the α_{ij} with $\alpha_i = k_\Theta(i)$. Let $f(i) = |\alpha_i|^{2r}$, and let ϕ denote the Fourier transform of f. Then

$$\phi(0) = \sum_i f(i) = \sum_i |\alpha_{ij}|^{2r} > 0. \tag{3.4.111}$$

Since f is the $2r$-fold product of αs, ϕ can be expressed as a $2r$-fold convolution product. Let $\beta = \chi^{-1}\mathcal{I}_\Theta$, and observe that β is the Fourier transform of α. One has

$$\begin{aligned}
\phi(0) &= \int \ldots \int \beta(\omega_1)\beta(\omega_2-\omega_1)\ldots\beta(\omega_{2r-1}-\omega_{2r-2})\beta(\omega_{2r-1})\,d\omega_1\ldots d\omega_{2r-1} \\
&= \int \beta(\omega_1)\,d\omega_1 \int \beta(\omega_2-\omega_1)\,d\omega_2 \ldots \int \beta(\omega_{2r-1}-\omega_{2r-2})\beta(\omega_{2r-1})\,d\omega_{2r-1} \\
&\le \{\int \beta(\omega)\,d\omega\}^{2(r-1)}\{\int \beta(\omega_{2r-1}-\omega_{2r-2})^2\,d\omega_{2r-1}\}^{1/2}\{\int \beta(\omega)^2\,d\omega\}^{1/2} \\
&= \{\int_\Theta \chi(\omega)^{-1}\,d\omega\}^{2(r-1)}\{\int_\Theta \chi(\omega)^{-2}\,d\omega\} \\
&= O\{\rho^{d(2r-1)}\}. \qquad (3.4.112)
\end{aligned}$$

Here, the inequality follows by applying Hölder's inequality. The last two lines above follow from the definition of β and Lemma 3.4.2.

Combining (3.4.111) and (3.4.112) gives (3.4.110). Substitution of (3.4.110) into (3.4.108) yields

$$\sum_i |\alpha_{ij}|^{s_m} = O\{\rho^{d(s_m-1)}\} \text{ for } s_m = 2,\ldots,2b.$$

The $2b$-th moments of (3.4.108) are therefore bounded by

$$\begin{aligned}
\mathbf{E}|\sum_{i\in\mathcal{R};i\neq j} \alpha_{ij}\epsilon_i\epsilon_j|^{2b} &= \max_{p=1,\ldots,b}\left[\prod_{m=1}^{p} O\{\rho^{d(s_m-1)}\}\right] \\
&= \max_{p=1,\ldots,b} O\{\rho^{d(2b-p)}\} \\
&= O(\rho^{bd}),
\end{aligned}$$

since $\sum_{m=1}^{p} s_m = 2b$, and $\rho \to 0$ as $n \to \infty$. The required result follows now immediately, since $\sum_{j\in\mathcal{R}} 1 = O(n^d)$. □

Lemma 3.4.11 *For $n > 0$, $\rho \in I = [k_1\delta_0,\ k_2\delta_0]$ put*

$$W = \sum_{j\in\mathcal{R}} (\epsilon \star k_{\Omega\setminus\Theta_\rho})(j)\{(h\star t)(j) - h(j)\}.$$

If $b \in \mathbb{N}$ and $\mathbf{E}\epsilon^{2b} < \infty$, then, as $n \to \infty$,

$$\mathbf{E}W^{2b} = O\{n^{bd}(n\rho)^{b(d-2a)}(\lambda\rho)^{4b}\}.$$

Proof of Lemma 3.4.11

Fix $n > 0$, $\rho \in I$. Put $\Theta = \Theta_\rho$, $r_{ij} = k_{\Omega\setminus\Theta}(i-j)$, $s_j = (h\star t)(j) - h(j)$. Then for $b \in \mathbb{N}$,

$$\mathbf{E}W^{2b} = \mathbf{E}(\sum_{j\in\mathcal{R}}\sum_l \epsilon_l r_{jl} s_j)^{2b}$$

$$
\begin{aligned}
&\le \sum_{j_1\in\mathcal{R}}\cdots\sum_{j_{2b}\in\mathcal{R}}\sum_{l_1}\cdots\sum_{l_{2b}}\mathbf{E}(\epsilon_{l_1}\ldots\epsilon_{l_{2b}}r_{j_1l_1}s_{j_1}\ldots r_{j_{2b}l_{2b}}s_{j_{2b}}) \\
&\le \sum_{\alpha=1}^{b} C_\alpha \sum_{l_1}\cdots\sum_{l_\alpha}\sum_{j_1\in\mathcal{R}} r_{j_1l_1}s_{j_1}\cdots\sum_{j_{2b}\in\mathcal{R}} r_{j_{2b}l_\alpha}s_{j_{2b}} \\
&\le \sum_{\alpha=1}^{b} C_\alpha \sum_{l_1}\{\sum_{j\in\mathcal{R}} r_{jl_1}s_j\}^{p_1}\sum_{l_2}\{\sum_{j\in\mathcal{R}} r_{jl_2}s_j\}^{p_2}\ldots\sum_{l_\alpha}\{\sum_{j\in\mathcal{R}} r_{jl_\alpha}s_j\}^{p_\alpha}, \qquad (3.4.113)
\end{aligned}
$$

where

$$
2 \le p_i \le 2b, \qquad \sum_{i=1}^{\alpha} p_i = 2b, \qquad (3.4.114)
$$

and C_α are positive constants into which the moments have been absorbed.

Replacing each term $\sum_{j\in\mathcal{R}} r_{jl}s_j$ by $\sum_j r_{jl}s_j$, and ignoring the constants C_α, one obtains

$$
\begin{aligned}
W_\infty &\equiv \sum_{\alpha=1}^{b}\sum_{l_1}\{\sum_j r_{jl_1}s_j\}^{p_1}\ldots\sum_{l_\alpha}\{\sum_j r_{jl_\alpha}s_j\}^{p_\alpha} \\
&= \sum_{\alpha=1}^{b}\sum_{l_1}(r\star s)(l_1)^{p_1}\ldots\sum_{l_\alpha}(r\star s)(l_\alpha)^{p_\alpha}. \qquad (3.4.115)
\end{aligned}
$$

Now,

$$
\sum_l (r\star s)(l)^p \begin{cases} = \sum_l (r\star s)(l)^{2q} & \text{if } p = 2q \\ \le \{\sum_l (r\star s)(l)^{4q}\sum_l (r\star s)(l)^2\}^{1/2} & \text{if } p = 2q+1, \end{cases} \qquad (3.4.116)
$$

where the inequality follows from Hölder's inequality. Let $g = (r\star s)^{2q}$ and let γ denote the Fourier transform of g. One has

$$
\sum_l (r\star s)(l)^{2q} = \sum_l g(l) = \gamma(0) > 0. \qquad (3.4.117)
$$

On the other hand, since g is the $2q$-fold product of the functions $r\star s$, γ is the $2q$-fold convolution product of ρ, where ρ denotes the Fourier transform of $r\star s$. Therefore $\gamma(0)$ may be written as

$$
\begin{aligned}
\gamma(0) &= \int\cdots\int \phi(\omega_1)\phi(\omega_2-\omega_1)\ldots\phi(\omega_{2q-1}-\omega_{2q-2})\phi(\omega_{2q-1})\, d\omega_1\ldots d\omega_{2q-1} \\
&\le c_1\{\int \phi(\omega)\, d\omega\}^{2q-2}\{\int \phi(\omega)^2\, d\omega\} \text{ for } c_1 > 0, \qquad (3.4.118)
\end{aligned}
$$

where we have used the fact that

$$
\begin{aligned}
|\int \alpha(x-y)\beta(y)\, dy| &= |\alpha\star\beta(x)| = c_2|\int a(\omega)b(\omega)e^{-i\langle\omega,x\rangle}\, d\omega| \\
&\le c_2\int |a(\omega)b(\omega)|\, d\omega = c_3\int |\alpha(x)\beta(x)|\, dx
\end{aligned}
$$

for Fourier transform pairs (a,α) and (b,β) and constants $c_2, c_3 > 0$.

Since ϕ denotes the Fourier transform of $r \star s$ (for a definition of r and s see the beginning of this proof),

$$\phi = \tau(1-\chi^{-1})\mathcal{I}_{\Omega\setminus\Theta}$$

and (3.4.118) can now be estimated as follows, using Lemma 3.4.6:

$$\begin{aligned}
\gamma(0) &\leq c_4\Big[\int_{\Omega\setminus\Theta} \tau(\theta)\{1-\chi(\theta)^{-1}\}\, d\theta\Big]^{2q-2} \int_{\Omega\setminus\Theta} \tau(\theta)^2\{1-\chi(\theta)^{-1}\}^2\, d\theta \\
&\leq c_5\{(n\rho)^{d-a}(\lambda\rho)^2\}^{2q-2}\{n^d(n\rho)^{d-2a}(\lambda\rho)^4\} \\
&= c_5 n^{p(d-a)}\rho^{(p-1)d-pa}(\lambda\rho)^{2p} \quad (c_4, c_5 > 0). \qquad (3.4.119)
\end{aligned}$$

By a similar argument to that given in (3.4.117)-(3.4.119), one obtains the following bound for the second estimate (i.e. the case $p = 2q+1$ in (3.4.116)):

$$\begin{aligned}
\{\sum_l (r\star s)(l)^{4q}\sum_l (r\star s)(l)^2\}^{1/2} &\leq c_6\{n^d(n\rho)^{(4q-1)d-4qa}(\lambda\rho)^{8q}\}^{1/2}\{n^d(n\rho)^{d-2a}(\lambda\rho)^4\}^{1/2} \\
&= c_6 n^{p(d-a)}\rho^{(p-1)d-pa}(\lambda\rho)^{2p} \quad (c_6 > 0). \qquad (3.4.120)
\end{aligned}$$

Substitution of (3.4.119) and (3.4.120) into (3.4.115) yields

$$\begin{aligned}
W_\infty &\leq c_7 \sum_{\alpha=1}^{b}\prod_{i=1}^{\alpha} n^{p_i(d-a)}\rho^{(p_i-1)d-p_ia}(\lambda\rho)^{2p_i} \\
&= c_7\sum_{\alpha=1}^{b} n^{2b(d-a)}\rho^{(2b-\alpha)d-2ba}(\lambda\rho)^{4b} \\
&= \max_{1\leq\alpha\leq b} O\{n^{2b(d-a)}\rho^{(2b-\alpha)d-2ba}(\lambda\rho)^{4b}\} \\
&= O\{n^{2b(d-a)}\rho^{bd-2ba}(\lambda\rho)^{4b}\}, \qquad (3.4.121)
\end{aligned}$$

since $\max_{1\leq\alpha\leq b}\rho^{-\alpha d} = \rho^{-bd}$.

This is the desired result for W_∞. To obtain the same bound for $\mathbf{E}W^{2b}$, one proceeds as in that part of the proof of Theorem 3.5 which concerns the term W, in order to show that $\sum_l(\sum_{j\notin\mathcal{R}} r_{jl}s_j)^p = o\{\sum_l(r\star s)(l)^p\}$ for $2 \leq p \leq 2b$. This then completes the proof of Lemma 3.4.11. □

Proof of Theorem 3.8

We are only concerned with the first cross-validation method here, and thus write CV instead of CV_1 in this proof. It will be necessary to regard CV, SSE and MSSE explicitly as functions of n; we do this by writing CV(n), SSE(n) and MSSE(n) as in (3.2.14).

Let $\rho_i = k_1\delta_0 + in^{-\gamma}$, $\phi_{i,r} = \rho_i + rn^{-\gamma}$ as in (3.3.16). If, as $n\to\infty$,

$$\sup_{0\leq i<\kappa}\sup_{0<r\leq 1}\left|\frac{\mathrm{CV}(n,\rho_i)-\mathrm{SSE}(n,\rho_i)}{\mathrm{MSSE}(n,\rho_i)} - \frac{\mathrm{CV}(n,\phi_{i,r})-\mathrm{SSE}(n,\phi_{i,r})}{\mathrm{MSSE}(n,\phi_{i,r})}\right| = O(n^{-1}) \quad a.s.;$$

and

$$\sup_{0\le i<\kappa}\left|\frac{\mathrm{CV}(n,\rho_i)-\mathrm{SSE}(n,\rho_i)}{\mathrm{MSSE}(n,\rho_i)}\right|\to 0 \quad a.s.,$$

then

$$\sup_{\delta\in I}\left|\frac{\mathrm{CV}(n,\delta)-\mathrm{SSE}(n,\delta)}{\mathrm{MSSE}(n,\delta)}\right|\to 0 \quad a.s. \text{ as } n\to\infty.$$

In Proposition 3.7, the first part was proved, and it therefore suffices to show that the second condition holds. For this, we use the following strategy. Put

$$\mathcal{S}_n = \sup_{0\le i<\kappa}\left|\frac{\mathrm{CV}(n,\rho_i)-\mathrm{SSE}(n,\rho_i)}{\mathrm{MSSE}(n,\rho_i)}\right|. \tag{3.4.122}$$

By the Borel-Cantelli lemma (see Theorem 4.2.1 and 4.2.2 of Chung (1974)), $\mathcal{S}_n\to 0$ $a.s.$ if for $\xi>0$

$$\sum_{n=1}^{\infty}\mathbf{P}(\mathcal{S}_n>\xi)<\infty. \tag{3.4.123}$$

Now,

$$\begin{aligned}
\sum_{n=1}^{\infty}\mathbf{P}(\mathcal{S}_n>\xi) &\le \sum_{n=1}^{\infty}\sum_{0\le i<\kappa}\mathbf{P}\Big\{\left|\frac{\mathrm{CV}(n,\rho_i)-\mathrm{SSE}(n,\rho_i)}{\mathrm{MSSE}(n,\rho_i)}\right|>\xi\Big\}\\
&\le \sum_{n=1}^{\infty}\sum_{0\le i<\kappa}\xi^{-p}\mathbf{E}\Big\{\left|\frac{\mathrm{CV}(n,\rho_i)-\mathrm{SSE}(n,\rho_i)}{\mathrm{MSSE}(n,\rho_i)}\right|^p\Big\} \quad (p>0)\\
&\le \xi^{-p}\sum_{n=1}^{\infty}\{\min_{0\le i<\kappa}\mathrm{MSSE}(n,\rho_i)\}^{-p}\sum_{0\le i<\kappa}\mathbf{E}|\mathrm{CV}(n,\rho_i)-\mathrm{SSE}(n,\rho_i)|^p\\
&\equiv \xi^{-p}\mathcal{E}.
\end{aligned} \tag{3.4.124}$$

In this calculation, the second last inequality follows from Chebyshev's inequality or Markov's inequality. Therefore

$$\xi^{-p}\mathcal{E}<\infty \text{ implies that } \mathcal{S}_n\to 0 \quad a.s.,$$

and thus the theorem follows, if we show that $\xi^{-p}\mathcal{E}<\infty$.

Estimates for $\mathrm{MSSE}(n,\rho_i)$ can be derived from Proposition 3.1. We now begin by calculating the p-th moments. Fix $n\in\mathbb{N}$. As in (3.4.26) and (3.4.76), put

$$\mathrm{CV}-\mathrm{SSE} = -2(U+V+W),$$

where

$$\begin{aligned}
U &= \sum_{j\in\mathcal{R}}\{\hat{t}^*(j)-\hat{t}(j)\}(h\star t)(j)\\
&= \sum_{j\in\mathcal{R}}\{(h\star t)(j)\mathrm{D}(h\star t)(j)+(h\star t)(j)\mathrm{D}\epsilon_j\}k_\Theta(0);\\
V &= \sum_{j\in\mathcal{R}}\{\hat{t}^*(j)-t(j)\}\epsilon_j
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j\in\mathcal{R}}[\{\mathcal{B}(j)+k_\Theta(0)\mathsf{D}(h\star t)(j)\}\epsilon_j+\{(k_\Theta\star\epsilon)(j)+k_\Theta(0)\mathsf{D}\epsilon_j\}\epsilon_j]; \quad \text{and} \\
W &= \sum_{j\in\mathcal{R}}\{\hat{t}(j)-t(j)-(\epsilon\star h^{(-1)})(j)\}\{(h\star t)(j)-t(j)\} \\
&= \sum_{j\in\mathcal{R}}\{\mathcal{B}(j)-(\epsilon\star h^{(-1)})(j)\}\{(h\star t)(j)-t(j)\}. \qquad (3.4.125)
\end{aligned}
$$

Then

$$
\mathbf{E}|\mathrm{CV}-\mathrm{SSE}|^p \le c_1\{\mathbf{E}|U|^p+\mathbf{E}|V|^p+\mathbf{E}|W|^p\} \quad (c_1>0). \qquad (3.4.126)
$$

To estimate the random parts in CV−SSE, we use Rosenthal's inequality (see (3.5.11) in the Appendix) in the following form: writing $S=\sum_{j\in\mathcal{R}}s_j$, where S is a martingale with independent increments, one gets

$$
\mathbf{E}|S|^{2b} \le c_2[\{\sum_{j\in\mathcal{R}}\mathbf{E}|s_j|^2\}^b+\sum_{j\in\mathcal{R}}\mathbf{E}|s_j|^{2b}] \quad (c_2>0). \qquad (3.4.127)
$$

Let $0\le i<\kappa$. Put $\rho=\rho_i$. Consider

$$
U_1=\sum_{j\in\mathcal{R}}(h\star t)(j)\mathsf{D}(h\star t)(j)k_\Theta(0),
$$

the deterministic part of U, as in (3.4.28), with

$$
\begin{aligned}
|\mathsf{D}(h\star t)(j)| &= |(2d)^{-1}\sum_{|k|=1}(h\star t)(j+k)-(h\star t)(j)| \\
&= (2d)^{-1}|\sum_{|k|=1}\sum_{l}h(l)\{t(j+k-l)-t(j-l)\}| \\
&\le (2d)^{-1}\sup_{|k|=1}|t(j+k)-t(j)|\sum_{l}h(l) \quad (j\in\mathbb{Z}^d) \\
&= O(n^{-\Delta d}), \qquad (3.4.128)
\end{aligned}
$$

since $\sum_l h(l)=1$, and since t satisfies the Lipschitz condition **A5**. Replacing $\sum_{j\in\mathcal{R}}\mathsf{D}(h\star t)(j)=o(n^d)$ in the proof of Theorem 3.5 by $\sum_{j\in\mathcal{R}}\mathsf{D}(h\star t)(j)=O(n^{d-\Delta d})$, in analogy with (3.4.30) it then follows that

$$
|U_1|^{2b}=O(n^{2bd-2b\Delta d}\rho^{2bd}). \qquad (3.4.129)
$$

For

$$
U_2=\sum_{j\in\mathcal{R}}(h\star t)(j)\mathsf{D}\epsilon_j k_\Theta(0),
$$

the random part of U, it follows from (3.5.2) and (3.5.3) that U_2 is a martingale with respect to the σ-field which was denoted by $\mathcal{F}_1$ in the Appendix. We may thus use

Rosenthal's inequality (3.4.127) to get

$$\begin{aligned}\sum_{j\in\mathcal{R}} \mathbf{E}|(h\star t)(j)\mathrm{D}\epsilon_j k_\Theta(0)|^2 &\le c_3 k_\Theta(0)^2 \sup_{j\in\mathcal{R}}|(h\star t)(j)|^2 \sum_{j\in\mathcal{R}}\mathbf{E}\epsilon_j^2 \\ &= O(n^d\rho^{2d}) \qquad (3.4.130)\end{aligned}$$

as in (3.4.31), for $c_3 > 0$. For the second term on the right hand side of (3.4.127), observe that

$$\begin{aligned}\sum_{j\in\mathcal{R}} \mathbf{E}|(h\star t)(j)\mathrm{D}\epsilon_j k_\Theta(0)|^{2b} &\le c_4 k_\Theta(0)^{2b}\sum_{j\in\mathcal{R}}(h\star t)(j)^{2b}\mathbf{E}\epsilon_j^{2b} \quad (c_4>0) \\ &= O(n^d\rho^{2bd}) \qquad (3.4.131)\end{aligned}$$

follows again by the same arguments as (3.4.31), since $\mathbf{E}\epsilon^{2b} < \infty$. Rosenthal's inequality for U_2, using (3.4.130) and (3.4.131) now yields

$$\mathbf{E}|U_2|^{2b} = O(n^{bd}\rho^{2bd}). \qquad (3.4.132)$$

Next consider

$$V_1 = \sum_{j\in\mathcal{R}}\{\mathcal{B}(j) + k_\Theta(0)\mathrm{D}(h\star t)(j)\}\epsilon_j,$$

the first part of V, as in (3.4.34). Since the ϵ_j are independent,

$$\sum_{j\in\mathcal{R}}\{\mathcal{B}(j)+k_\Theta(0)\mathrm{D}(h\star t)(j)\}^2\mathbf{E}\epsilon_j^2 = O\{n^d(n\rho)^{d-2a}\} + o(n^d\rho^{2d}) \qquad (3.4.133)$$

as in (3.4.36). On the other hand,

$$\begin{aligned}\sum_{j\in\mathcal{R}}\mathbf{E}[\{\mathcal{B}(j)+k_\Theta(0)\mathrm{D}(h\star t)(j)\}\epsilon_j]^{2b} &\le c_5\sum_{j\in\mathcal{R}}\{\mathcal{B}(j)^{2b} + k_\Theta(0)^{2b}\mathrm{D}(h\star t)(j)^{2b}\} \\ &= O\{n^d(n\rho)^{2b(d-a)}\} + o(n^d\rho^{2bd}) \qquad (3.4.134)\end{aligned}$$

for $c_5 > 0$ follows from Lemmas 3.4.1 and 3.4.2.

Substitution of (3.4.133) and (3.4.134) into Rosenthal's inequality leads to

$$\mathbf{E}|V_1|^{2b} = O\{n^{bd}(n\rho)^{b(d-2a)} + n^d(n\rho)^{2b(d-a)}\} + o(n^{bd}\rho^{2bd}). \qquad (3.4.135)$$

Now turn to

$$\begin{aligned}V_{21} &= (2d)^{-1}k_\Theta(0)\sum_{|l|=1}\sum_{j\in\mathcal{R}}\epsilon_j\epsilon_{j+l} \\ &= d^{-1}k_\Theta(0)\sum_{l\in\mathcal{P}}\sum_{j\in\mathcal{R}}\epsilon_j\epsilon_{j+l},\end{aligned}$$

where $\mathcal{P} = \{l \in \mathbb{Z}^d \; : \; |l| = 1 \text{ and } l_i \leq 0 \; \forall i = 1, \ldots, d\}$. It follows that

$$\begin{aligned} \mathbf{E}(\sum_{l\in\mathcal{P}} \epsilon_j\epsilon_{j+l})^2 &\leq c_6\sigma^4 \quad \text{and} \\ \mathbf{E}(\sum_{l\in\mathcal{P}} \epsilon_j\epsilon_{j+l})^{2b} &= \sum_{l_1\in\mathcal{P}} \cdots \sum_{l_{2b}\in\mathcal{P}} \mathbf{E}(\epsilon_j^{2b}\epsilon_{j+l_1}\ldots\epsilon_{j+l_{2b}}) \leq c_7, \end{aligned} \tag{3.4.136}$$

for $c_6, c_7 > 0$ and $j \in \mathbb{Z}^d$. From the Appendix it follows that V_{12} is a martingale with respect to the σ-field $\mathcal{F}_2'$ defined in (3.5.6), and therefore (3.4.127) can be applied:

$$\begin{aligned} \mathbf{E}|V_{21}|^{2b} &\leq c_8 k_\Theta(0)^{2b}\{(\sum_{j\in\mathcal{R}} 1)^b + (\sum_{j\in\mathcal{R}} 1)\} \quad (c_8 > 0) \\ &= O(n^{bd}\rho^{2bd}) \end{aligned} \tag{3.4.137}$$

by Lemma 3.4.2 and (3.4.136).

Let V_{22} denote the remaining term in V (see (3.4.125)). Then

$$V_{22} = \sum_{j\in\mathcal{R}} \sum_{l\neq j} k_\Theta(j-l)\epsilon_j\epsilon_l.$$

Let

$$v_m = \sum_{l\in\mathcal{R}_m} \sum_{j\in\mathcal{R}; j\prec l} k_\Theta(j-l)\epsilon_j\epsilon_l \quad \text{for } m \in \mathbb{N}.$$

By (3.5.4) and (3.5.5), v_m is a martingale with respect to the σ-field which was denoted by $\mathcal{F}_2$ in the Appendix, and we now have to find bounds for the moments of the increments, in order to apply Rosenthal's inequality. From (3.4.39) it follows that

$$\sum_{j\in\mathcal{R}} \mathbf{E}\{\sum_{l\in\mathcal{R}_m; l\neq j} k_\Theta(j-l)\epsilon_j\epsilon_l\}^2 = O(n^d\rho^d),$$

while

$$\sum_{j\in\mathcal{R}} \mathbf{E}\{\sum_{l\in\mathcal{R}_m; l\neq j} k_\Theta(j-l)\epsilon_j\epsilon_l\}^{2b} = O(n^d\rho^{bd}),$$

by Lemma 3.4.10. Together these two estimates lead to

$$\mathbf{E}|v_m|^{2b} = O(n^{bd}\rho^{bd}). \tag{3.4.138}$$

To obtain a corresponding estimate for V_{22} instead of v_m, note that

$$\mathbf{E}|V_{22}|^{2b} \leq c_9\{\mathbf{E}|V_{22} - v_m|^{2b} + \mathbf{E}|v_m|^{2b}\} \quad (c_9 > 0). \tag{3.4.139}$$

Now,

$$\mathbf{E}|V_{22} - v_m|^{2b} = \mathbf{E}|\sum_{j\in\mathcal{R}} \sum_{l\notin\mathcal{R}_m; l\neq j} k_\Theta(j-l)\epsilon_j\epsilon_l|^{2b}$$

$$\begin{aligned} &\leq c_{10} \Big| \sum_{j\in\mathcal{R}} \sum_{l\notin\mathcal{R}_m} k_\Theta(j-l)^2 \Big|^b \quad (c_{10} > 0) \\ &= o(n^{bd}\rho^{bd}), \end{aligned} \tag{3.4.140}$$

by Lemma 3.4.4. The step involving the constant c_{10} can be obtained by an argument similar to that given in the proof of Lemma 3.4.10. Substitution of (3.4.140) and (3.4.138) into (3.4.139) now yields

$$\mathbf{E}|V_{22}|^{2b} = O(n^{bd}\rho^{bd}). \tag{3.4.141}$$

We now turn to W and put

$$W_1 = \sum_{j\in\mathcal{R}} \mathcal{B}(j)\{(h\star t)(j) - t(j)\}.$$

Since W_1 is purely deterministic, (3.4.46) can be used to give the following bound

$$|W_1|^{2b} = O\{n^{2bd}(n\rho)^{2b(d-2a)}(\lambda\rho)^{4b}\}. \tag{3.4.142}$$

For

$$W_2 = \sum_{j\in\mathcal{R}} (\epsilon\star k_\Theta)(j)\{(h\star t)(j) - t(j)\},$$

the random part of W, Lemma 3.4.11 provides the following bound:

$$\mathbf{E}|W_2|^{2b} = O\{n^{bd}(n\rho)^{b(d-2a)}(\lambda\rho)^{4b}\}. \tag{3.4.143}$$

Having calculated bounds for U, V, W, we now return to (3.4.126). Substitution of (3.4.129), (3.4.132), (3.4.135), (3.4.137), (3.4.141)–(3.4.143) into (3.4.126) leads to

$$\begin{aligned} \mathbf{E}|\mathrm{CV} - \mathrm{SSE}|^{2b} \leq\ & c_{11}\{(n\rho)^{2bd}n^{-2b\Delta d} + n^{bd}(n\rho)^{b(d-2a)} + n^d(n\rho)^{2b(d-a)} + (n\rho)^{bd} \\ & + n^{2bd}(n\rho)^{2b(d-2a)}(\lambda\rho)^{4b} + n^{bd}(n\rho)^{b(d-2a)}(\lambda\rho)^{4b}\} \end{aligned}$$

for $c_{11} > 0$. Now, if $\rho = k_1\delta_0 + in^{-\gamma}$ for some $0 \leq i < \kappa$, then $\rho \asymp \delta_0 = n^{-1+d/2a}$. Furthermore, $\lambda \leq c_{12}n^{1-\eta-d/2a}$ for $c_{12} > 0$ by assumption **A4**, and thus we get

$$\begin{aligned} \mathbf{E}|\mathrm{CV} - \mathrm{SSE}|^{2b} &\leq c_{13}(n^{-2b\Delta d + bd^2/a} + n^{bd^2/2a} + n^{-4b\eta + bd^2/a}) \\ &= c_{13}n^{bd^2/a}(n^{-2b\Delta d} + n^{-4b\eta} + n^{-bd^2/a}), \end{aligned}$$

from which it follows that

$$\begin{aligned} \sum_{0\leq i<\kappa} \mathbf{E}|\mathrm{CV} - \mathrm{SSE}|^{2b} &\leq \kappa\mathbf{E}|\mathrm{CV} - \mathrm{SSE}|^{2b} \\ &\leq c_{14}n^{\gamma-1+d/2a+bd^2/a}(n^{-2b\Delta d} + n^{-4b\eta} + n^{-bd^2/a}), \end{aligned} \tag{3.4.144}$$

since $\kappa \leq k_3n^{\gamma-1+d/2a}$.

From Proposition 3.1, Corollary 3.2 (or from Lemma 3.4.9 which precedes the proof of Proposition 3.7) it follows that $\mathrm{MSSE}(n,\rho_i) \asymp n^{d^2/2a}$, and therefore

$$m_n \equiv \{\min_{0\le i<\kappa} \mathrm{MSSE}(n,\rho_i)\}^{-2b} \asymp n^{-bd^2/a}. \tag{3.4.145}$$

Combining (3.4.144) and (3.4.145) gives rise to the following estimate for (3.4.124):

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbf{P}(\mathcal{S}_n > \xi) &\le c_{15}\sum_{n=1}^{\infty} n^{-bd^2/a} n^{\gamma-1+d/2a+bd^2/a}(n^{-2b\Delta d} + n^{-4b\eta} + n^{-bd^2/a}) \\
&= c_{15}\sum_{n=1}^{\infty} n^{\gamma-1+d/2a}(n^{-2b\Delta d} + n^{-4b\eta} + n^{-bd^2/a}) \\
&\le c_{15}\sum_{n=1}^{\infty} n^{-1-\omega} < \infty,
\end{aligned} \tag{3.4.146}$$

since $b > (\gamma+\omega+d/2a)\max\{2a/d^2,\ 1/4\eta\}$. Equation (3.4.146) shows that $\sum_{n=1}^{\infty} \mathbf{P}(\mathcal{S}_n > \xi)$ is finite, and therefore, by the Borel-Cantelli lemma (see (3.4.122) and (3.4.123)),

$$\sup_{0\le i<\kappa} \left| \frac{\mathrm{CV}(n,\rho_i) - \mathrm{SSE}(n,\rho_i)}{\mathrm{MSSE}(n,\rho_i)} \right| \to 0 \quad a.s. \text{ as } n \to \infty$$

as required. This completes the proof of Theorem 3.8. □

3.4.8 Proof of Proposition 3.9

Proposition 3.9 *Assume that t, h and ϵ satisfy* **A1–A3** *and that $\lambda = O(n^{1-d/2a})$. Assume further that $a > 3d$ and $\gamma > 2 + 2a - d/2a$. If*

1. $b \ge (\gamma + \omega + d/2a)2a/d$ *for $\omega > 0$, and*

2. $\mathbf{E}|\epsilon|^r < \infty$ *for $1 \le r \le 2b$,*

then

$$\sup_{\delta\in I} \left| \frac{SSE(n,\delta) - MSSE(n,\delta)}{MSSE(n,\delta)} \right| \to 0 \quad a.s. \quad as\ n \to \infty.$$

Outline of Proof of Proposition 3.9

The proof of this proposition is based on the steps and results obtained in the proof of Proposition 3.4. It can be carried out along the same lines as described in the proofs of Proposition 3.7 and Theorem 3.8. For this reason, we shall not go into any details here, but just indicate the estimates and bounds for the various key steps. The notation used here is that established in the proofs of Propositions 3.4 and 3.7, and Theorem 3.8.

We begin by mimicking the results of Proposition 3.7, that is, we make use of the inequality

$$\left|\frac{\mathcal{D}(x)}{\mathcal{M}(x)} - \frac{\mathcal{D}(y)}{\mathcal{M}(y)}\right| \leq |\mathcal{D}(x) - \mathcal{D}(y)||\mathcal{M}(x)|^{-1} + |\mathcal{D}(y)||\mathcal{M}(x)\mathcal{M}(y)|^{-1}|\mathcal{M}(x) - \mathcal{M}(y)|. \tag{3.4.147}$$

Here $\mathcal{M}$ will denote MSSE and $\mathcal{D}$ the difference SSE – MSSE. We use the notation of (3.4.14), and let $\mathcal{D} = 2S + T_1 + T_2$.

Put $y = \rho_i$, $x = \phi_{i,r}$ with $0 \leq i < \kappa$ and $0 < r \leq 1$. Put

$$\begin{aligned} \mathcal{B}_{x\setminus y}(j) &= (2\pi)^{-d} \int_{\Theta_x \setminus \Theta_y} \tau(\theta) e^{-i\langle j,\theta\rangle}\, d\theta; \\ k_{x\setminus y}(j) &= (2\pi)^{-d} \int_{\Theta_x \setminus \Theta_y} \chi(\theta)^{-1} e^{-i\langle j,\theta\rangle}\, d\theta, \end{aligned}$$

and let $\mathcal{B}_x$ (respectively k_x) be defined as in (3.4.74).

One obtains the following bounds for the terms on the right hand side of (3.4.147). For

$$S = \sum_{j\in\mathcal{R}} \mathcal{B}(j)(k_\Theta \star \epsilon)(j)$$

as in (3.4.13), one has

$$\begin{aligned} \mathbf{E}\{S(x) - S(y)\}^2 &\leq c_1\sigma^2 \sum_l [\{\sum_{j\in\mathcal{R}} \mathcal{B}_x(j) k_{x\setminus y}(j-l)\}^2 + \{\sum_{j\in\mathcal{R}} \mathcal{B}_{x\setminus y}(j) k_y(j-l)\}^2] \\ &= o[n^{3d-2a}\{x^{d-2a}|x-y|^d + |x^{d-2a} - y^{d-2a} y^d\}] \end{aligned}$$

in analogy with (3.4.16)–(3.4.20). Applying the results of Lemma 3.4.8, it follows that

$$\begin{aligned} S(x) - S(y) &= o\{n^{d^2/4a+(1-\gamma)d/2} + n^{d^2/2a-d/4a+(1-\gamma)/2}\} \quad a.s.; \\ S(y) &= o(n^{d^2/2a}) \quad a.s. \end{aligned} \tag{3.4.148}$$

For

$$T_1 = \sum_l (\epsilon_l^2 - \sigma^2) \sum_{j\in\mathcal{R}} k_\Theta(j-l)^2,$$

$$\begin{aligned} \mathbf{E}\{T_1(x) - T_1(y)\}^2 &= \sum_l (\mathbf{E}\epsilon_l^4 - \sigma^4)\{\sum_{j\in\mathcal{R}} k_x(l-j)^2 - k_y(l-j)^2\}^2 \\ &= O(|x-y|)^d, \end{aligned}$$

and therefore

$$\begin{aligned} T_1(x) - T_1(y) &= O(|x-y|^{d/2}) = O(n^{-\gamma d/2}) \quad a.s.; \\ T_1(y) &= O(n^{d^2/4a-d/2}) \quad a.s. \end{aligned} \tag{3.4.149}$$

follows from (3.4.22)–(3.4.24), Parseval's identity and Lemmas 3.4.2 and 3.4.8.

Lastly, putting

$$T_2 = \sum_i \sum_{l \neq i} \epsilon_i \epsilon_l \sum_{j \in \mathcal{R}} k_\Theta(j-i) k_\Theta(j-l),$$

yields

$$\begin{aligned}
\mathbf{E}\{T_2(x) - T_2(y)\}^2 &= \sum_i \sum_{l \neq i} \sum_m \sum_{r \neq m} \mathbf{E}(\epsilon_i \epsilon_l \epsilon_m \epsilon_r) \\
&\quad \times \sum_{j_1 \in \mathcal{R}} \sum_{j_2 \in \mathcal{R}} \{k_x(j_1 - i) k_{x \setminus y}(j_1 - l) k_x(j_2 - m) k_{x \setminus y}(j_2 - r) \\
&\qquad + k_{x \setminus y}(j_1 - i) k_y(j_1 - l) k_{x \setminus y}(j_2 - m) k_y(j_2 - r)\} \\
&= O[n^d \{(x^{d/2} + y^{d/2}) |x-y|^{d/2} + |x-y|^d\}]
\end{aligned}$$

by a calculation similar to that of (3.4.25). It follows that

$$\begin{aligned}
T_2(x) - T_2(y) &= O\{n^{(1+d/2a-\gamma)d/4} + n^{(1-d)\gamma/2}\} \quad a.s.; \\
T_2(y) &= O(n^{d^2/4a}) \quad a.s.,
\end{aligned} \tag{3.4.150}$$

again by Lemmas 3.4.2 and 3.4.8 and Parseval's identity.

Using the bounds for $\mathcal{M}$ given in Lemma 3.4.9 and substituting (3.4.148)–(3.4.150) into (3.4.147) leads to

$$\left| \frac{\mathcal{D}(x)}{\mathcal{M}(x)} - \frac{\mathcal{D}(y)}{\mathcal{M}(y)} \right| = o(n^{-1}) \quad a.s., \tag{3.4.151}$$

since $\gamma > 2 + 2a - d/2a$.

So far, we have sketched the analogue of the proof of Proposition 3.7. It remains to find bounds for the $2b$-th moments in order to apply the Borel-Cantelli lemma. As in the proof of Theorem 3.8, one can make use of Rosenthal's inequality in the form discussed in the Appendix, since S, T_1 and T_2 are martingales with respect to the σ-fields denoted in the Appendix by $\mathcal{F}_1$ and $\mathcal{F}_2$. One obtains the following bounds:

$$\begin{aligned}
\mathbf{E}S^{2b} &= O\{n^{bd}(n\rho)^{2b(d-a)-b}\}, \\
\mathbf{E}T_1^{2b} &= O(n^{bd}\rho^{2bd}), \quad \text{and} \\
\mathbf{E}T_2^{2b} &= O(n^{bd}\rho^{bd}),
\end{aligned}$$

from which one concludes that

$$\begin{aligned}
\mathbf{E}\mathcal{D}(n, \rho_i)^{2b} &\leq c_2 n^{bd}\{(n\rho_i)^{2b(d-a)-b} + \rho_i^{bd}\} \\
&\leq c_3 n^{bd^2/a - bd/2a}
\end{aligned}$$

for $c_2, c_3 > 0$.

Taking into account the results of Lemma 3.4.9, we argue as follows:

$$\begin{aligned}
&\sum_{n=1}^{\infty} \mathcal{M}(n)^{-2b} \sum_{0\leq i<\kappa} \mathbf{E}\mathcal{D}(n,\rho_i)^{2b} \\
\leq\ & c_4 \sum_{n=1}^{\infty} n^{-bd^2/a} n^{\gamma-1+(d/2a)+(bd^2/a)-(bd/2a)} \\
<\ & c_4 \sum_{n=1}^{\infty} n^{-1-w} \quad (c_4>0),
\end{aligned}$$

since $b > (\gamma + \omega + d/2a)2a/d$.

The desired result follows from Markov's inequality and the Borel-Cantelli lemma as in the proof of Theorem 3.8. □

3.5 Appendix: Rosenthal's Inequality

For $d \geq 2$, $\mathbb{Z}^d$ does not have a natural ordering like $\mathbb{Z}$; however, one can define an ordering in many different ways. For our purpose it is convenient to make a particular choice, denoted by $\prec$ with the following properties. Define the n-shell B_n of $\mathbb{Z}^d$ to be

$$B_n = \{j \in \mathbb{Z}^d \ : \ \|j\|_\infty = n\}.$$

For each n-shell choose a linear order, say $<_n$. For $j, k \in \mathbb{Z}^d$ define $j \prec k$ as follows.

1. If j and k belong to the same shell B_n, say, then

$$j \prec k \ \text{ if and only if } \ j <_n k.$$

2. If $j \in B_n$ and $k \in B_m$ for $n \neq m$, then

$$j \prec k \ \text{ if and only if } \ n < m.$$

(The second property corresponds to ordering subsets according to their sup-norm $\|.\|_\infty$.) For $j \in \mathbb{Z}^d$, let j^- denote the predecessor of j with respect to $\prec$. Then

$$j^- = \max_k \{k \prec j\}. \tag{3.5.1}$$

In what follows we describe two types of martingales and give the appropriate form of Rosenthal's inequality. For more details on martingales see Hall and Heyde (1980).

We begin with the definition of a martingale. Let I be an ordered set in $\mathbb{Z}$. Let $\{\mathcal{F}_i \ : \ i \in I\}$ be an increasing sequence of σ-fields. A sequence of random variables

$\{Z_i \ : \ i \in I\}$ is called a *martingale* with respect to $\{\mathcal{F}_i\}$ if

1. Z_i is measurable with respect to $\mathcal{F}_i$;
2. $\mathbf{E}|Z_i| < \infty$;
3. $\mathbf{E}(Z_i|\mathcal{F}_j) = Z_j \quad a.s.$ for all $j < i$.

For our purpose, the crucial property of the index set I in the above definition is the fact that it is ordered. Since we want to define martingales indexed by $j \in \mathbb{Z}^d$, we shall consider subsets of $\mathbb{Z}^d$ with the $\prec$-ordering, such as the regions $\mathcal{R}_n$ defined in the text in Chapter 3. Let $\{\epsilon_j \ : \ j \in \mathbb{Z}^d\}$ denote independent random variables with zero mean, finite variance and finite absolute moments. For $j \in \mathbb{Z}^d$ let

$$\mathcal{F}_{1,j} = \sigma\{\epsilon_k \ : \ k \preceq j\} \tag{3.5.2}$$

denote the σ-field generated by the ϵ_k, $k \preceq j$. For $\alpha \in l^1(\mathbb{Z}^d)$, $j \in \mathbb{Z}^d$, put

$$S_{1,j} = \sum_{k \preceq j} \alpha_k \epsilon_k. \tag{3.5.3}$$

It follows immediately that $S_{1,j}$ is an $\mathcal{F}_{1,j}$-martingale.

For the second example, take $\{\epsilon_j \ : \ j \in \mathbb{Z}^d\}$ as in (3.5.2). For $j \in \mathbb{Z}^d$ define

$$\mathcal{F}_{2,j} = \sigma\{\epsilon_k \epsilon_l \ : \ l \prec k \preceq j\} \tag{3.5.4}$$

to be the σ-algebra generated by pairs $\epsilon_k \epsilon_l$. Take $j \in \mathbb{Z}^d$, $\alpha \in l^1(\mathbb{Z}^d \times \mathbb{Z}^d)$ and put

$$S_{2,j} = \sum_{k \preceq j} s_k \quad \text{with } s_k = \sum_{l \prec k} \alpha_{kl} \epsilon_k \epsilon_l. \tag{3.5.5}$$

Note that $S_{2,j}$ is an $\mathcal{F}_{2,j}$-martingale, since

$$\begin{aligned} \mathbf{E}(S_{2,j}|\mathcal{F}_{2,i}) &= \Big(\sum_{k \preceq i} + \sum_{i \prec k \prec j}\Big)\Big\{\sum_{l \prec k} \alpha_{kl} \mathbf{E}(\epsilon_k \epsilon_l | \mathcal{F}_{2,i})\Big\} \\ &= \sum_{k \preceq i} \sum_{l \prec k} \alpha_{kl} \epsilon_k \epsilon_l = S_{2,i} \end{aligned}$$

for $i \prec j$, since $\mathbf{E}(\epsilon_k \epsilon_l | \mathcal{F}_{2,i}) = \mathbf{E}(\epsilon_k \epsilon_l) = 0$ for $i \prec k \preceq j$.

As a special case of the S_2 martingale defined above, put

$$\mathcal{F}'_{2,j} = \sigma\{\epsilon_k \epsilon_{k+l} \ : \ k \preceq j, \ l \in \mathcal{P}\} \tag{3.5.6}$$

where $\mathcal{P} = \{l \in \mathbb{Z}^d \ : \ |l| = 1, \ l_i \leq 0 \ \forall i = 1, \ldots, d\}$, and let

$$S'_{2,j} = \sum_{k \preceq j} s'_k, \qquad s'_k = \sum_{l \in \mathcal{P}} \epsilon_k \epsilon_{k+l}. \tag{3.5.7}$$

Clearly $S'_{2,j}$ is an $\mathcal{F}'_{2,j}$-martingale.

Before we present the form of Rosenthal's inequality which is appropriate for S_1 and S_2, observe that S_1 and S_2 are both sums of independent random increments, and thus

$$V_{1,j}^2 \equiv \sum_{k \preceq j} \mathbf{E}(\alpha_k^2 \epsilon_k^2 | \mathcal{F}_{k^-}) = \sigma^2 \sum_{k \preceq j} \alpha_k^2; \tag{3.5.8}$$

$$\begin{aligned} V_{2,j}^2 &\equiv \sum_{k \preceq j} \mathbf{E}(s_k^2 | \mathcal{F}_{k^-}) = \sum_{k \preceq j} \sum_{l \prec k} \sum_{m \prec k} \alpha_{kl} \alpha_{km} \mathbf{E}(\epsilon_k \epsilon_l \epsilon_k \epsilon_m | \mathcal{F}_{k^-}) \\ &= \sum_{k \preceq j} \sum_{l \prec k} \alpha_{kl}^2 \mathbf{E}\epsilon_k^2 \mathbf{E}\epsilon_l^2 = \sigma^4 \sum_{k \preceq j} \sum_{l \prec k} \alpha_{kl}^2. \end{aligned} \tag{3.5.9}$$

Here k^- denotes the predecessor of k as defined in (3.5.1).

The full version of Rosenthal's inequality can be found in Theorem 2.12, p23 of Hall and Heyde (1980). We are only concerned with the upper bound part for the martingales S_1 and S_2 here. The proof of Rosenthal's inequality for our index sets follows from that given in Hall and Heyde (1980) by essentially replacing their indexing $1 \leq i \leq n$ by our ordered indexing sets in $\mathbb{Z}^d$. We shall therefore not give a proof here.

Because of the independence of the increments in S_1 and S_2, and because of (3.5.8) and (3.5.9), it follows that

$$\begin{aligned} \mathbf{E}[\{\sum_{k \preceq j} \mathbf{E}(\alpha_k^2 \epsilon_k^2 | \mathcal{F}_{k^-})\}^r] &= \mathbf{E}(V_{1,j}^2)^r = (V_{1,j}^2)^r \\ \mathbf{E}[\{\sum_{k \preceq j} \mathbf{E}(s_k^2 | \mathcal{F}_{k^-})\}^r] &= \mathbf{E}(V_{2,j}^2)^r = (V_{2,j}^2)^r. \end{aligned} \tag{3.5.10}$$

Hence, if we write X_k for $\alpha_k \epsilon_k$ or for s_k, for $S_j = \sum_{k \preceq j} X_k$ the upper bound of Rosenthal's inequality reduces to:

For $1 \leq r < \infty$ there exists a constant $c > 0$ such that

$$\mathbf{E}|S_j|^{2r} \leq c[\{\sum_{k \preceq j} \mathbf{E}(X_k^2 | \mathcal{F}_{k^-})\}^r + \sum_{k \preceq j} \mathbf{E}|X_k|^{2r}], \tag{3.5.11}$$

where $\mathcal{F}_k = \mathcal{F}_{1,k}$ or $\mathcal{F}_k = \mathcal{F}_{2,k}$ as appropriate.

Chapter 4

Median Smoothing

4.1 Introduction

In linear regression the least squares method is a very popular estimation procedure. It is linear and thus 'nice' from a mathematical viewpoint—and is also relatively easy to compute. For the normal model it furthermore results in a minimum variance unbiased linear estimator. However for data with outliers, or for models other than the normal, the least squares estimator does not perform so well. For this reason, one often turns to more robust estimation methods which may not be quite as efficient as least squares under the normal model, but which will perform well in a broad range of situations. We shall concentrate on one such method, the least absolute deviation method, which leads us to consider a running median estimator, rather than a running mean estimator. This estimator is in many ways more robust than the running mean. In fact for estimation of a location parameter the median is the most robust estimator with respect to the robustness notions associated with Hampel's 'influence function' and Rousseeuw and Hampel's 'change-of-variance function' (see Section 2.5c of Hampel *et al.* (1986)).

Apart from its importance as a robust estimator, the median has its own place in image processing: median filtering is a commonly used technique for detecting and preserving edges and for filtering out impulses (see Gallagher and Wise (1981), Yang and Huang (1981) and Bovik *et al.* (1987)). Many interesting algorithms have been derived for the median, as well as some more theoretical properties. The continuing interest in the median as an edge-preserving tool can be seen in the growing number of research papers in this area (see e.g. Huang (1981), Chin and Yeh (1983), Bovik *et al.* (1983) and (1987), and references therein).

It may also be of interest to observe that the edge-preserving property of the median has a mathematical counterpart in the 'local-shift sensitivity parameter λ', which is infinite in the case of the median (see p88ff of Hampel *et al.* (1986)).

Motivated by the robustness properties as well as by its widespread use in image analysis, we shall focus here on the asymptotic behaviour of the running median or median smoother. For this, we use the framework of M-estimation and robust smoothing methods which has been described briefly in Section 1.5 of Chapter 1. Insert opposite page

For regression models of the form

$$Y_i = m(x_i) + \epsilon_i \quad \text{for } i = 1, \dots, n, \tag{4.1.1}$$

Priestley and Chao (1972) suggested the use of the linear estimator m_n^* given by

$$m_n^*(x) = \sum_{i=1}^{n} \alpha_i(x) Y_i$$

for weights $\alpha_i(x)$ derived from a kernel density function (see (1.4.5)). Estimates of this form are frequently used, partly because of their flexibility and smoothing properties. They are certainly appropriate if the ϵ_i are normally distributed, but for other noise distributions, Härdle and Gasser (1984) have suggested the use of the weighted M-estimator approach: The unknown function m in (4.1.1) is estimated by m_n, where $m_n(x)$ is a zero of the function H_n given by

$$H_n(x, \cdot) = \sum_{i=1}^{n} \alpha_i(x) \psi(Y_i - \cdot) \tag{4.1.2}$$

for a suitably chosen function ψ.

The method proposed by Härdle and Gasser presupposes that the function ψ of (4.1.2) has a bounded derivative ψ' and that $\psi'(0)$ is positive. We shall refer to M-estimators of this type as *smooth* M-estimators. As we shall see in the next section, the median smoother is an M-estimator but not a smooth M-estimator. We also deviate from Härdle and Gasser in that their regression functions are defined on $[-1, 1]$, while we consider d-dimensional domains ($d \geq 1$) for our images. This generalisation to higher dimensions affects the bias in a rather surprising way, with $d = 4$ the dimension at which both the terms that contribute to bias are of the same order (see Theorem 4.4).

In the approach we adopt here, we choose a family of smooth M-estimators which converge to the median smoother. Thus instead of considering a single estimator and its asymptotic behaviour as the sample size increases, we deal with a family of M-estimators simultaneously. Our results describe the asymptotic mean square error as the sample size increases and as the M-estimators approximate the median smoother, and from this we are able to determine the asymptotic mean square error of the median smoother. The standard with which we compare our results is the rate of convergence of the mean smoother. To obtain this rate of convergence for the median smoother as well, the rate of convergence of the M-estimators to the median has to be chosen correctly: if the rate is too slow, bias increases too much.

The chapter is organised in the following way: in Section 4.2 we describe our model, and the family of M-estimators approximating the median. In Section 4.3 we first present the results for our approximating M-estimators (Propositions 4.2 and 4.3) and then in Theorem 4.4 we derive the rate of convergence for the mean square error of the median. Corresponding results for the mean estimator are also briefly described in this section. Proofs of our results are deferred to Section 4.4.

4.2 Image Models and the Median Smoother

4.2.1 Models for the Observations

In picture transmission or picture scanning by television cameras, the input picture is usually distorted by additive random noise which is independent of the input. As in these examples, our observations in this chapter will consist of true images which are degraded by random noise. Such observations can also be regarded as coming from d-dimensional ($d \geq 1$) regression models.

A. The true image T. We assume that the true image is a deterministic real-valued function T which is defined on a region in $\mathbf{R}^d$, $d \geq 1$. For mathematical convenience we consider the compact region $J^d = [-1, 1]^d$.

We assume that $\nabla^2 T$ exists and is bounded. In this chapter we let ∇ denote the derivative with respect to $x \in \mathbf{R}^d$; thus for any g defined on $\mathbf{R}^d$

$$\begin{aligned} \nabla g &= (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d})^T g \\ \nabla^2 g &= (\frac{\partial^2}{\partial x_i \partial x_j}) g \quad \text{for } i, j = 1, \ldots, d. \end{aligned} \tag{4.2.1}$$

We say $\nabla^2 T$ is *bounded* if $\sup_{\|x\| \leq 1} |\langle x, \nabla^2 T x \rangle| < \infty$.

B. The random noise ϵ. Let ϵ denote a random function defined on J^d, and assume that for $x \in J^d$, $\epsilon_x = \epsilon(x)$ is a random variable with probability density function $f_x = f(\cdot; x)$. This defines a function $f : \mathbf{R} \times J^d \to \mathbf{R}$. For convenience we sometimes refer to f as the probability density of the random function ϵ. By abuse of notation, but in conformity with the derivative notation ∇ defined in (4.2.1), we write ∇f instead of $\nabla f(\epsilon; \cdot)$. Thus, ∇f refers to differentiation with respect to the second argument $x \in J^d$, but not with respect to the first. For each random variable ϵ_x, let

$$\mathbf{E}\epsilon_x = \int \varepsilon f(\varepsilon; x) \, d\varepsilon. \tag{4.2.2}$$

The use of a continuously defined noise model allows us to make assumptions about the way the density of ϵ svaries spatially as the grid becomes finer. It can also result in different noise processes for each realisation. See also (4.2.3) below.

We asume that ϵ and f satisfy the following:

N1 The function $f(\cdot;x)$ is symmetric about 0, and $f(0;x) > 0$ for $x \in J^d$;

N2 The derivative $\nabla^2 f$ exists, and ∇f, $\nabla^2 f$ are absolutely integrable;

N3 The function $f(\cdot;x)$ satisfies a Lipschitz condition at 0 for $x \in J^d$;

N4 The ϵ_x are independent with mean zero.

C. The observed data Y. For $x \in J^d$, T and ϵ as in the preceding two paragraphs, put

$$Y(x) = T(x) + \epsilon_x. \tag{4.2.3}$$

Equation (4.2.3) defines a random function $Y : J^d \to \mathbf{R}$, and each random variable $Y(x)$ has an induced probability density function $f_Y(\cdot;x)$ given by

$$f_Y(y;x) = f\{y - T(x);x\}, \tag{4.2.4}$$

where $f(\cdot;x)$ is the probability density function of ϵ_x. The density f_Y is twice differentiable with respect to $x \in \mathbf{R}^d$. This follows from **N2** above. Furthermore,

$$\mathbf{E}Y(x) = \int y f_Y(y;x)\,dy = \int \{T(x) + \varepsilon\} f(\varepsilon;x)\,d\varepsilon = T(x) \tag{4.2.5}$$

follows from (4.2.2) and **N4**.

The true image T is defined everywhere in J^d, but observations of T can only be obtained at discrete points in J^d. For $n > 0$, define $\mathcal{G}_n \subseteq \mathbb{Z}^d$ by

$$\mathcal{G}_n = \{j = \{j^{(i)}\} \in \mathbb{Z}^d \ : \ |j^{(i)}| \le n,\ i = 1,\ldots,d\} \tag{4.2.6}$$

(see also paragraph A of Subsection 3.2.1). The grid $\mathcal{G}_n$ is regular, square-based and consists of $O(n^d)$ equally spaced points and $\mathcal{G}_n = \mathcal{R}_n$ with $K = 1$ in the definition of $\mathcal{R}_n$ (see (3.2.13)). The notation for the components $j^{(i)}$ of $j \in \mathbb{Z}^d$ differs from the subscript notation of Chapter 2 to avoid confusion with the notation used for the sampling points x_j in (4.2.7), below. With each $j \in \mathcal{G}_n$ we associate a sampling point $x_j \in J^d$, given by

$$x_j = n^{-1} j, \tag{4.2.7}$$

and then consider data of the form

$$Y_j = T(x_j) + \epsilon_j, \quad \text{for } j \in \mathcal{G}_n. \tag{4.2.8}$$

A comparison with (4.2.3) shows that $Y_j = Y(x_j)$ and $\epsilon_j = \epsilon_{x_j}$. This slight ambiguity in the notation for ϵ will not be a problem and is outweighed by the simpler notation of (4.2.8).

4.2.2 The Median Smoother $\widehat{T}$

To estimate T from the data Y_j, we use the *median smoother* or running median, denoted by $\widehat{T}$, which is defined as follows:

For $x \in J^d$, put

$$U(x,\cdot) = \sum_{j\in\mathcal{G}_n} \alpha_j(x)|Y_j - \cdot|, \tag{4.2.9}$$

for weights $0 \le \alpha_j(x) \le 1$, such that $\sum_{j\in\mathcal{G}_n} \alpha_j(x) \le 1$, and let

$$\widehat{T}(x) = \arg\min U(x,\cdot). \tag{4.2.10}$$

Sometimes $\widehat{T}(x)$ is called the LAD-estimator (least absolute deviation estimator) as in Bloomfield and Steiger (1983).

To allow for edge effects, we require only $\sum_{j\in\mathcal{G}_n} \alpha_j(x) \le 1$, and not $\sum_{j\in\mathcal{G}_n} \alpha_j(x) = 1$. Since we are concerned with asymptotic behaviour here, the choice of method for dealing with edge effects is not very important.

We choose the weights in the following way. Fix $n > 0$. For $x = \{x^{(i)}\} \in J^d$ define d-dimensional cubes V_x of volume n^{-d} by

$$V_x = \left\{ z = \{z^{(i)}\} \in J^d \;:\; x^{(i)} - \frac{1}{2n} \le z^{(i)} < x^{(i)} + \frac{1}{2n}, \quad \forall i = 1,\dots,d \right\}. \tag{4.2.11}$$

For $0 < k < n$, put $h = (2k+1)/(2n)$. For $x \in J^d$, $j \in \mathcal{G}_n$, define $\alpha_j(x)$ by

$$\alpha_j(x) = \begin{cases} (2k+1)^{-d} & \text{if } V_{x_j - x} \cap [-h,h]^d \supseteq I_n \\ 0 & \text{otherwise} \end{cases} \tag{4.2.12}$$

where x_j is as in (4.2.7) and I_n denotes a closed cube of volume $(2n)^{-d}$. As can be seen, at most $(2k+1)^d$ weights are non-zero, and $\alpha_j(x) = 0$ if x_j is outside a certain window centred on x. Putting

$$L_k(x) = \{ j \in \mathcal{G}_n \;:\; \alpha_j(x) \ne 0 \} \tag{4.2.13}$$

leads to

$$U(x,\cdot) = \sum_{j\in L_k(x)} \alpha_j(x)|Y_j - \cdot| = (2k+1)^{-d} \sum_{j\in L_k(x)} |Y_j - \cdot|. \tag{4.2.14}$$

The weights $\alpha_j(x)$ defined in (4.2.12) are special cases of the weights derived from kernel functions: assume that $\kappa : J^d \to \mathbf{R}$ is symmetric and $\int_{J^d} \kappa(u)\, du = 1$. For $x \in J^d$, V_{x_j} as in (4.2.11) for some $x_j \in J^d$, and $0 < h \le 1$, put

$$\alpha_j(x) = h^{-d} \int_{V_{x_j}} \kappa\left(\frac{u - x}{h}\right) du. \tag{4.2.15}$$

In this situation, h is often called the bandwidth or smoothing parameter and is regarded as a function of n, as is done in kernel density estimation or nonparametric regression. In these two research areas, one wants to choose h asymptotically optimally, where 'optimality' refers to the minimisation of an appropriately chosen measure of performance, such as those described in Subsection 1.4.1.

Our weights can be derived from (4.2.15) by putting

$$\kappa(u) = 2^{-d} \quad \text{and} \quad h = (2k+1)/(2n). \tag{4.2.16}$$

Unlike nonparametric regression or density estimation, the significant parameter for us here is k (instead of h), which determines the window size. ⊕ Insert opposite page Our aim is to select $k = k(n)$ asymptotically optimally (see the paragraph following (4.2.18)). Note that (4.2.12) reduces the 'effective' observations Y_j to those which satisfy $j \in L_k(x)$.

From (4.2.9)–(4.2.12), one can derive the fact that

$$\widehat{T}(x) = \widehat{T}(x_\ell) \quad \text{for } x \in V_{x_\ell}, \tag{4.2.17}$$

that is, $\widehat{T}$ is constant on the cubes V_{x_ℓ}, $\ell \in \mathcal{G}_n$. On the other hand, if $x = x_\ell$ for some $\ell \in \mathcal{G}_n$, and if $\hat{\tau}(x)$ denotes the median of the observations Y_j, $j \in L_k(x)$, then it follows immediately that $\hat{\tau}(x) = \arg\min U(x, \cdot)$. Thus it suffices to estimate $\widehat{T}$ at the grid-points x_j, $j \in L_k(x)$.

To evaluate the performance of the median smoother $\widehat{T}$, we consider the (pointwise) mean square error (MSE) of $\widehat{T}$. For $x \in J^d$, put

$$\text{MSE}\{\widehat{T}(x)\} = \mathbf{E}\{\widehat{T}(x) - T(x)\}^2 \tag{4.2.18}$$

(as in (2.4.3)).

The estimator $\widehat{T}$ clearly depends on n and k. For given n, how is k chosen? In the sequel we shall regard k as a function of n and then select that k which minimises the order of the mean square error (4.2.18) as $n \to \infty$. Such a k will be called *asymptotically optimal.*

4.2.3 Approximations to the Median Smoother

As mentioned in the introduction to this chapter, we employ methods used in the development of the theory of M-estimation in order to calculate the mean square error of $\widehat{T}$. These methods require differentiability properties which the median does not possess, and we therefore cannot adopt this approach directly.

We generalise the definition of M-estimator given in (1.5.2) to include weights. We call $\hat{\tau}$ a *weighted M-estimator* for the observations Y_j if

$$\hat{\tau} = \arg\min \sum a_j \rho(Y_j - \cdot) \tag{4.2.19}$$

for weights a_j and some 'distance' function ρ. Using this definition, it follows that the median smoother $\widehat{T}$ of (4.2.10) is a weighted M-estimator with $\rho_0(z) = |z|$ in place of ρ in (4.2.19).

The approach adopted here is to construct a family of convex functions $\{\rho_\nu \, : \, \nu > 0\}$ such that

1. $\rho_\nu \to \rho_0$ as $\nu \to 0$;
2. for each $\nu > 0$, ρ_ν is even and $\rho_\nu \in C^2(\mathbf{R})$.

The first property of the ρ_ν suffices to ensure that the M-estimators $\widehat{T}^\nu$ corresponding to ρ_ν converge to $\widehat{T}$. The second allows us to calculate the MSE for each $\widehat{T}^\nu$ in terms of expected value and variance of H^ν (see (4.1.2) and Section 1.5). Combining these two results and letting ν decrease at a suitable rate enables us to estimate the MSE of $\widehat{T}$ from the corresponding estimate for $\widehat{T}^\nu$.

For $\nu > 0$ define functions $\rho_\nu : \mathbf{R} \to \mathbf{R}$ by

$$\rho_\nu(z) = \{z^2 + \nu^2\}^{1/2}. \tag{4.2.20}$$

For $n > 0$ and observations Y_j of the form (4.2.8), put

$$U^\nu(x, \cdot) = \sum_{j \in L_k(x)} \alpha_j(x) \rho_\nu(Y_j - \cdot), \tag{4.2.21}$$

where the weights $\alpha_j(x)$ and the sets $L_k(x)$ are those of (4.2.12) and (4.2.13), respectively. From (4.2.20) it follows that

$$\lim_{\nu \to 0} \rho_\nu(z) = \rho_0(z).$$

The convexity of U^ν implies that for $x \in J^d$

$$\{\tau^\nu \in \mathbf{R} \, : \, \tau^\nu = \arg\min U^\nu(x, \cdot)\}$$

is non-empty, convex and compact. This result follows as in Lemma 1 of Huber (1964) and shows that a minimum of U^ν exists. For $x \in J^d$ we put

$$\widehat{T}^\nu(x) = \arg\min U^\nu(x, \cdot). \tag{4.2.22}$$

Clearly $\widehat{T}^\nu$ is a smooth M-estimator corresponding to the convex C^2-function ρ_ν. In the next section we consider some of the properties of $\widehat{T}^\nu$, which will then be used in the MSE calculations for $\widehat{T}$.

4.3 Results

For the results given in this section we regard the window size parameter k and the parameter ν of the family of M-estimators $\widehat{T}^{\nu}$ as functions of n. For notational convenience we let **L1–L4** denote the statements below about the asymptotic behaviour of k and ν as $n \to \infty$:

L1 $k(n) \to \infty$;

L2 $k(n)/n \to 0$;

L3 $\nu(n) \to 0$;

L4 $\nu(n) = O\{k^{-d}n^{-1}\max(k^2n^{-1}, 1)\}$.

The parameters k and n refer to the window size and sample size in each dimension. Sometimes it will be advantageous to use the parameters K and N given by

$$K = (2k+1)^d, \quad N = (2n+1)^d. \tag{4.3.1}$$

We call K the *effective sample size*, and N the *sample size*, and we regard K as a function of N.

For the convenience of the reader we briefly summarise properties of the image T, the noise ϵ and the probability density function f of ϵ, which were described in the previous section. For $n > 0$, consider data Y of the form

$$Y_j = T(x_j) + \epsilon_j, \quad \text{for } j \in \mathcal{G}_n, \tag{4.3.2}$$

where

A1 the true image $T : J^d \to \mathbf{R}$ possesses a bounded second derivative $\nabla^2 T$;

A2 the ϵ_j are independent with mean zero;

A3 the density f has first and second absolutely integrable derivatives ∇f and $\nabla^2 f$;

A4 the family $\{f(\cdot; x) \ : \ x \in J^d\}$ is symmetric about 0, satisfies a Lipschitz condition at 0 and is strictly positive at 0.

4.3.1 Convergence of $\widehat{T}^{\nu}$ to $\widehat{T}$

We begin by proving the convergence of the estimators $\widehat{T}^{\nu}$ to the median $\widehat{T}$. As can easily be seen from the definition of U and U^{ν} (see (4.2.14) and (4.2.21)),

$$U^{\nu}(x, \omega) \to U(x, \omega) \quad \text{as } \nu \to 0, \text{ for } x \in J^d, \omega \in \mathbf{R},$$

but it remains to be shown that the minimiser of the U^ν converges to that of U as $\nu \to 0$. (Note that U is defined on an odd number of points and therefore has a unique minimiser.) In the asymptotic results below, we let $k \to \infty$ (as $n \to \infty$). To indicate the dependence of $\widehat{T}$, $\widehat{T}^\nu$, U and U^ν on k, we write for $x \in J^d$

$$\begin{aligned}\widehat{T}_k(x) &= \arg\min U_k(x,\cdot)\\ \widehat{T}_k^\nu(x) &= \arg\min U_k^\nu(x,\cdot)\end{aligned} \tag{4.3.3}$$

where

$$\begin{aligned}U_k(x,\cdot) &= \sum_{j\in L_k(x)} \alpha_j(x)|Y_j-\cdot|\\ U_k^\nu(x,\cdot) &= \sum_{j\in L_k(x)} \alpha_j(x)\{(Y_j-\cdot)^2+\nu^2\}^{1/2}.\end{aligned} \tag{4.3.4}$$

With this notation we obtain the following relationship between $\widehat{T}_k^\nu$ and $\widehat{T}_k$.

Proposition 4.1 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L3**. *If $x \in J^d$, and if $\nu \le k^{-d}n^{-1}\max\{k^2n^{-1},1\}$, then*

$$|\widehat{T}_k^\nu(x) - \widehat{T}_k(x)| = O\left\{(\frac{k}{n})^2 + n^{-1}\right\} \quad \textit{as } n \to \infty.$$

The proposition shows that the $\widehat{T}_k^\nu$ approximate $\widehat{T}_k$ as $\nu \to 0$. In fact, it tells us explicitly how to choose ν, here as a function of k as well as of n, in order to guarantee that $\widehat{T}_k^\nu(x)$ is within an $O\{(\frac{k}{n})^2+n^{-1}\}$ neighbourhood of $\widehat{T}_k(x)$, a fact which we rely on in Theorem 4.4.

A proof of Proposition 4.1 can be found in Subsection 4.4.1.

4.3.2 The Mean Square Error of the $\widehat{T}^\nu$

We begin with some notation. For $z \in \mathbf{R}$, put

$$\begin{aligned}\psi_\nu(z) &:= z\{z^2+\nu^2\}^{-1/2}\\ \psi_\nu'(z) &= \nu^2\{z^2+\nu^2\}^{-3/2}.\end{aligned} \tag{4.3.5}$$

For data Y_j, $j \in \mathcal{G}_n$, as in (4.3.2), and for $x \in J^d$, define

$$\begin{aligned}H_k^\nu(x,z) &= \sum_{j\in L_k(x)} \alpha_j(x)\psi_\nu(Y_j-z)\\ D_k^\nu(x,z) &= \sum_{j\in L_k(x)} \alpha_j(x)\psi_\nu'(Y_j-z).\end{aligned} \tag{4.3.6}$$

Recall that if $\rho_\nu(z) = \{z^2 + \nu^2\}^{1/2}$, as in (4.2.20), then

$$\begin{aligned}\psi_\nu(z) &= \frac{d}{dz}\rho_\nu(z),\\ \psi'_\nu(z) &= \frac{d^2}{dz^2}\rho_\nu(z).\end{aligned}$$

An immediate consequence of this relationship is that

$$\widehat{T}_k^\nu(x) = \arg\min U_k^\nu(x,\cdot)$$

implies

$$H_k^\nu\{x, \widehat{T}_k^\nu(x)\} = 0. \tag{4.3.8}$$

For each $x \in J^d$, the function H_k^ν is differentiable with respect to the second variable and we may thus write

$$H_k^\nu\{x, \widehat{T}_k^\nu(x)\} = H_k^\nu\{x, T(x)\} + \tilde{D}_k^\nu\{x, T(x)\}\{T(x) - \widehat{T}_k^\nu(x)\}. \tag{4.3.9}$$

Here we have used a Taylor expansion of each summand of H_k^ν about $T(x)$, the true image at x, and

$$\tilde{D}_k^\nu\{x, T(x)\} = \sum \alpha_j(x)\psi'_\nu\{Y_j - T(x) + \eta_j\} \tag{4.3.10}$$

denotes the remainder with $\eta_j = \theta_j\{T(x) - \widehat{T}_k^\nu(x)\}$, for some θ_j, $0 < \theta_j < 1$. Since $\widehat{T}_k^\nu(x)$ is a root of $H_k^\nu\{x,\cdot\}$, we obtain the following expression

$$\widehat{T}_k^\nu(x) - T(x) = H_k^\nu\{x, T(x)\}\left[\tilde{D}_k^\nu\{x, T(x)\}\right]^{-1}. \tag{4.3.11}$$

If one shows that $\tilde{D}_k^\nu\{x, T(x)\}$ converges to some non-random $\gamma_\nu(x)$, then it will follow that

$$\mathbf{E}\{\widehat{T}_k^\nu(x) - T(x)\}^2 = \mathbf{E}H_k^\nu\{x, \widehat{T}_k^\nu(x)\}^2\gamma_\nu(x)^{-2},$$

that is, the mean square error of $\widehat{T}_k^\nu$ can be expressed in terms of $\mathbf{E}H_k^\nu\{x, \widehat{T}_k^\nu(x)\}^2$.

This is the approach we adopt here. The idea goes back at least as far as Cramér's proof of the asymptotic normality of the maximum likelihood estimator (see Section 33.3 of Cramér (1946)).

We begin with estimates for moments of H_k^ν. For $x \in J^d$, $k \in \mathbf{N}$, $\nu > 0$, put

$$\begin{aligned}\beta_k^\nu &= \mathbf{E}H_k^\nu\{x, T(x)\}\\ v_k^\nu &= \operatorname{var} H_k^\nu\{x, T(x)\}\\ \gamma_\nu(x) &= \mathbf{E}\psi'_\nu(\epsilon_x).\end{aligned} \tag{4.3.12}$$

In this notation we have

Proposition 4.2 *Assume that T and ϵ satisfy* **A1–A4**, *and that k and ν satisfy* **L1–L3**.

If $x \in J^d$, then there exist $c_1, c_2 > 0$ such that

$$\beta_k^\nu \le c_1 f(0;x)\left\{(\frac{k}{n})^2 + n^{-1}\right\}\left\{1 + O(\nu^2) + O(\frac{k}{n})\right\}$$

and

$$v_k^\nu \le c_2 k^{-d}\left[1 + O\{(\frac{k}{n})^2\}\right].$$

Furthermore, there exists $c_3 > 0$ such that, as $n \to \infty$,

$$\mathbf{E}H_k^\nu\{x, T(x)\}^2 \le c_3\left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\}\left\{1 + O(\nu^2) + O(\frac{k}{n})\right\}.$$

The proof of this proposition is given in Subsection 4.4.2. The object of Proposition 4.2 was to calculate expectation and variance of H_k^ν. It now remains to obtain an estimate for $\tilde{D}_k^\nu$, in order to be able to derive estimates of the bias and variance of $\widehat{T}_k^\nu$ (see (4.3.11)). This is the concern of the next proposition.

Proposition 4.3 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L4**. *If $x \in J^d$, then*

$$MSE\{\widehat{T}_k^\nu(x)\} \le c_1\left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\}\{1 + o(1)\}$$

as $n \to \infty$, for some $c_1 > 0$.

A proof of Proposition 4.3 is given in Subsection 4.4.3. We delay a discussion of the above results until after our main result, Theorem 4.4.

4.3.3 The Mean Square Error of $\widehat{T}$

Our next result gives a general bound for MSE$\{\widehat{T}(x)\}$ as well as establishing optimal choices for the parameter k. As defined in Subsection 4.2.2, k is asymptotically optimal if it minimises the order of MSE.

To derive the mean square error of the median smoother $\widehat{T}(x)$ from the preceding results, we make use of the equality

$$\widehat{T}_k - T = (\widehat{T}_k - \widehat{T}_k^\nu) + (\widehat{T}_k^\nu - T). \tag{4.3.13}$$

In Proposition 4.1, the rate of convergence of the estimators $\widehat{T}_k^\nu$ to $\widehat{T}_k$ was given. This result together with the estimate of the mean square error of $\widehat{T}_k^\nu$ as given in Proposition 4.3 leads to

Theorem 4.4 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L4**. *If*

$x \in J^d$, then as $n \to \infty$,

$$MSE\{\widehat{T}_k(x)\} = O\left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\}. \tag{4.3.14}$$

Furthermore, optimal choices of k and the associated MSE are as follows:

1. *If $1 \leq d \leq 4$, then $k^\star(n) = n^{4/(4+d)}$ minimises the order of MSE and*

$$MSE\{\widehat{T}_{k^\star}(x)\} = O\{n^{-4d/(4+d)}\}. \tag{4.3.15}$$

2. *If $d \geq 4$, then $k^\star(n) = n^{2/d}$ minimises the order of MSE and*

$$MSE\{\widehat{T}_{k^\star}(x)\} = O(n^{-2}). \tag{4.3.16}$$

A proof of Theorem 4.4 can be found in Subsection 4.4.4. Note that the first two terms in (4.3.14) are due to squared bias and the term k^{-d} is due to the variance of $\widehat{T}_k(x)$. The optimal rates of convergence of MSE (4.3.15)–(4.3.16) may be rather surprising, in the sense that they seem to indicate that the rates of convergence of MSE do not decrease as the dimension increases. This is in fact not the case as the next corollary shows, in which we re-state the result in terms of the effective sample size K and the actual sample size N. Recall from (4.3.1) that $K = (2k+1)^d$ and $N = (2n+1)^d$.

Corollary 4.5 *Under the assumptions of Theorem 4.4,*

$$MSE\{\widehat{T}_k(x)\} = O\left\{(K/N)^{4/d} + N^{-2/d} + K^{-1}\right\}. \tag{4.3.17}$$

Furthermore, the optimal rates of convergence of MSE are as follows:

1. *If $1 \leq d \leq 4$, then $MSE\{\widehat{T}_{k^\star}(x)\} = O\{N^{-4/(4+d)}\}$.*
2. *If $d \geq 4$, then $MSE\{\widehat{T}_{k^\star}(x)\} = O(N^{-2/d})$.*

4.3.4 Discussion of Results

Comparison with the mean smoother. As indicated in the introduction of this chapter, we want to compare the rates of convergence of the median smoother with those of the mean smoother. For this, we impose assumptions **A1–A4** on the true image T and the error ϵ, and we take the weights $\alpha_j(x)$ as defined in (4.2.12). This implies that we consider the mean smoother at x on $(2k+1)^d$ points centred on $x \in J^d$.

Let $U_k^{(2)}$, $H_k^{(2)}$ and $D_k^{(2)}$ denote the analogues for the mean of U_k, H_k and D_k. Then for $x \in J^d$, $z \in \mathbf{R}$,

$$\begin{aligned} U_k^{(2)}(x,z) &= \sum_{j \in L_k(x)} \alpha_j(x)(Y_j - z)^2 \\ H_k^{(2)}(x,z) &= -2 \sum_{j \in L_k(x)} \alpha_j(x)(Y_j - z) \\ D_k^{(2)}(x,z) &= 2. \end{aligned} \tag{4.3.18}$$

If $\widehat{T}_k^{(2)}(x)$ denotes the minimiser of $U_k^{(2)}(x,\cdot)$, then, in analogy with (4.3.11), one obtains for $x \in J^d$

$$\widehat{T}_k^{(2)}(x) - T(x) = H_k^{(2)}\{x, T(x)\}\left[D_k^{(2)}\{x, T(x)\}\right]^{-1} \tag{4.3.19}$$

and therefore it follows immediately that

$$\mathrm{MSE}\{\widehat{T}_k^{(2)}(x)\} = \frac{1}{4}\mathbf{E}H_k^{(2)}\{x, T(x)\}^2.$$

Now, replacing ψ_ν by the identity function and ψ'_ν by the constant 2 in the proof of Proposition 4.2 simplifies that proof immensely and yields

$$\mathbf{E}H_k^{(2)}\{x, T(x)\} = (\frac{k}{n})^2 \int_{J^d} \kappa(u)\langle u, \nabla^2 T(x + \theta_3 hu)u\rangle \, du + O(n^{-1}) \quad (0 < \theta_3 < 1)$$

and

$$\mathrm{var}\ H_k^{(2)}\{x, T(x)\} = O(k^{-d}),$$

from which one concludes that

$$\mathrm{MSE}\{\widehat{T}_k^{(2)}(x)\} = O\left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\}, \tag{4.3.20}$$

as $n \to \infty$. A comparison of (4.3.20) with Theorem 4.4 shows that the median smoother performs as well asymptotically as the mean smoother.

Dependence of bias on dimension. Proposition 4.3, Theorem 4.4 and (4.3.20) show that, asymptotically, the M-estimators $\widehat{T}^\nu$, the median $\widehat{T}$ and the mean $\widehat{T}^{(2)}$ converge to the true image at the same rate (although the constants may of course be different). The term k^{-d} is due to variance of the estimators and reflects the fact that variance decreases as the effective sample size increases. In contrast to this, bias, which is $O\{(\frac{k}{n})^2 + n^{-1}\}$, does not seem to depend on the dimension of the problem. This behaviour may be due to the fact that we have used a regular square-based grid, and thus n^{-1} and $(\frac{k}{n})^2$ are the same in each direction. For other grids, such as triangular-based or hexagonal grids, it would be interesting to see whether bias behaves as in our case or whether bias reflects the worst one-dimensional direction.

Dependence of optimal rate on dimension. It is interesting to observe the effect of the dimension on the optimal rate of convergence, as given in Theorem 4.4 and

Corollary 4.5. One first observes that the optimal rate of convergence decreases as the dimension d increases. More surprising, however, is that one has to distinguish two cases, $1 \leq d \leq 4$ and $d \geq 4$, and that the optimal k and the corresponding rate of MSE are given by different functions in each case. The split into two cases is due to the order of bias: If $1 \leq d \leq 3$, the term in n^{-1} is negligible compared with that in $(\frac{k}{n})^2$. For $d = 4$, and for the optimal window size parameter k, the terms in $(\frac{k}{n})^2$ and n^{-1} are of the same order. When $d > 4$, the term in n^{-1} becomes the dominant term and the term in $(\frac{k}{n})^2$ becomes negligible. The importance and dominance of the grid spacing n^{-1} for higher-dimensional observations is rather unexpected in view of existing one-dimensional results in nonparametric regression, where this term is usually negligible (see p127-131 of Eubank (1988)).

Dependence of chosen ν on dimension. For the optimally chosen k, the chosen ν decreases more rapidly as d increases. However, for $d \geq 4$, one no longer requires $\nu(n)k(n)^{2-d}n^{-2} = O(1)$, and $\nu(n) = O(k^{-d}n^{-1})$ suffices, since it is no longer necessary that the rate of convergence of $\widehat{T}_k^\nu$ to $\widehat{T}$ be smaller than $(\frac{k}{n})^2$. (As we have seen in the previous paragraph, for $d \geq 4$, the bias of $\widehat{T}$ is $O(n^{-1})$.)

4.4 Proofs

In this section we give the proofs of Propositions 4.1–4.3 and of Theorem 4.4. For the ease of the reader, we repeat the statements of our results before proving them. Propositions and Theorems are numbered as in Section 4.3, while the lemmas are numbered consecutively as Lemma 4.4.1, 4.4.2, etc.

4.4.1 Proof of Proposition 4.1

Proposition 4.1 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L3**. *If $x \in J^d$, and if $\nu \leq k^{-d}n^{-1}\max\{k^2n^{-1}, 1\}$, then*

$$|\widehat{T}_k^\nu(x) - \widehat{T}_k(x)| = O\left\{(\frac{k}{n})^2 + n^{-1}\right\} \quad as\ n \to \infty.$$

Proof of Proposition 4.1

For $x \in J^d$, $\omega \in \mathbb{R}$, k and ν such that **L1–L3** hold, recall that

$$\begin{aligned} U_k^\nu(x,\omega) &= \sum_{j\in L_k(x)} \alpha_j(x)\rho_\nu(Y_j - \omega) \\ U_k(x,\omega) &= \sum_{j\in L_k(x)} \alpha_j(x)|Y_j - \omega|, \end{aligned}$$

where $\rho_\nu(z) = \{z^2 + \nu^2\}^{1/2}$. It follows that

$$\begin{aligned}\sup_k |U_k^\nu(x,\omega) - U_k(x,\omega)| &= \sup_k \sum_{j \in L_k(x)} \alpha_j(x) \left[\{(Y_j - \omega)^2 + \nu^2\}^{1/2} - |Y_j - \omega|\right] \\ &\leq \sup_k \sum_{j \in L_k(x)} \alpha_j(x) \{|Y_j - \omega| + \nu - |Y_j - \omega|\} \\ &\leq \nu.\end{aligned}$$

Furthermore, from the definitions of U_k^ν and U_k, one has

$$U_k^\nu(x,\omega) \geq U_k(x,\omega),$$

and therefore for $x \in J^d, \omega \in \mathbb{R}$,

$$U_k(x,\omega) \leq U_k^\nu(x,\omega) \leq U_k(x,\omega) + \nu. \tag{4.4.3}$$

Now let $\widehat{T}_k^\nu(x) = \arg\min U_k^\nu(x,\cdot)$ and $\widehat{T}_k(x) = \arg\min U_k(x,\cdot)$. Then

$$U_k^\nu\{x, \widehat{T}_k^\nu(x)\} \leq U_k^\nu\{x, \widehat{T}_k(x)\} \leq U_k\{x, \widehat{T}_k(x)\} + \nu$$

follows immediately from (4.4.3).

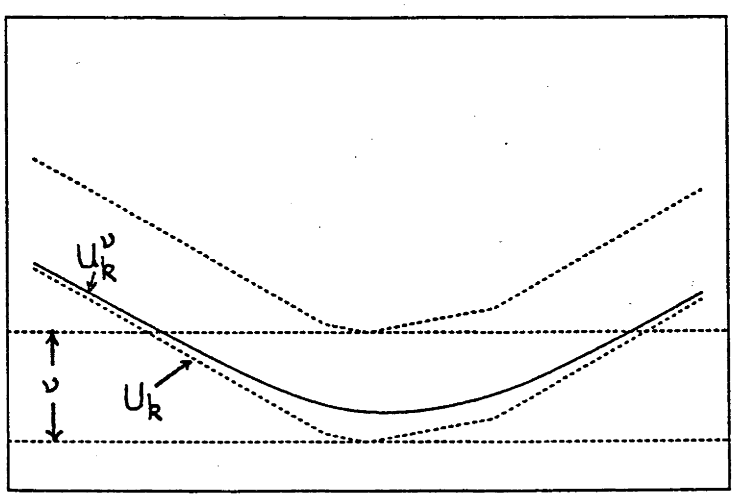

Figure 4.1: Location of minimisers of U_k and U_k^ν.

As Figure 4.4.1 shows,

$$|\widehat{T}_k^\nu(x) - \widehat{T}_k(x)| \leq \nu(2k+1)^d, \tag{4.4.4}$$

since $U_k^\nu\{x,\widehat{T}_k^\nu(x)\}$ belongs to the epigraph of U_k, $U_k^\nu\{x,\widehat{T}_k^\nu(x)\}$ is bounded above by $U_k\{x,\widehat{T}_k^\nu(x)\}+\nu$, and $(2k+1)^{-d}$ is a lower bound for the absolute value of the slope of the curve U_k. For k, $n \in \mathbb{N}$, let $\nu \le k^{-d}n^{-1}\max\{k^2n^{-1},1\}$. Equation (4.4.4) now leads to

$$|\widehat{T}_k^\nu(x)-\widehat{T}_k(x)| \le c_1 \max\{k^2n^{-2},n^{-1}\} \quad (c_1>0)$$

as required. □

4.4.2 Proof of Proposition 4.2

Proposition 4.2 *Assume that T and ϵ satisfy* **A1–A4**, *and that k and ν satisfy* **L1–L3**. *If $x \in J^d$, then there exist $c_1, c_2 > 0$ such that*

$$\beta_k^\nu \le c_1 f(0;x)\left\{(\frac{k}{n})^2+n^{-1}\right\}\left\{1+O(\nu^2)+O(\frac{k}{n})\right\}$$

and

$$v_k^\nu \le c_2 k^{-d}\left[1+O\{(\frac{k}{n})^2\}\right].$$

Furthermore, there exists $c_3>0$ such that, as $n \to \infty$,

$$\mathbf{E}H_k^\nu\{x,T(x)\}^2 \le c_3\left\{(\frac{k}{n})^4+n^{-2}+k^{-d}\right\}\left\{1+O(\nu^2)+O(\frac{k}{n})\right\}.$$

Before we prove the proposition, recall from (4.3.12) that for $x \in J^d$

$$\begin{aligned}\beta_k^\nu &= \mathbf{E}H_k^\nu\{x,T(x)\}\\ v_k^\nu &= \operatorname{var} H_k^\nu\{x,T(x)\}\\ \gamma_\nu(x) &= \mathbf{E}\psi'(\epsilon_x).\end{aligned}$$

We begin with Lemmas 4.4.1 and 4.4.2, which are used in the proof of Proposition 4.2.

Lemma 4.4.1 *Assume that k satisfies* **L1** *and* **L2** *and T and ϵ satisfy* **A1–A3**. *For $x \in J^d$, take $x_j \in J^d$ such that $j \in L_k(x)$. If $u \in V_{x_j}$ and $\nu>0$, then*

$$|\mathbf{E}\psi_\nu\{Y_j-T(x)\}-\mathbf{E}\psi_\nu\{Y(u)-T(x)\}| \le c_1\gamma_\nu(x)n^{-1}\{1+O(\frac{k}{n})\} \quad (c_1>0).$$

Proof of Lemma 4.4.1

Fix x, $x_j \in J^d$. For $u \in V_{x_j}$, consider

$$\begin{aligned}E &\equiv \mathbf{E}\psi_\nu\{Y_j-T(x)\}-\mathbf{E}\psi_\nu\{Y(u)-T(x)\}\\ &= \int \psi_\nu\{y-T(x)\}\{f_Y(y;x_j)-f_Y(y;u)\}\,dy\end{aligned}$$

$$= \; -\int \psi_\nu\{y - T(x)\}\langle u - x_j, \nabla f_Y(y; x_j + \xi_j)\rangle \; dy. \qquad (4.4.6)$$

Here we have used the Taylor expansion of f_Y about x_j, and $\xi_j = \theta_1(u - x_j)$ for some θ_1 with $0 < \theta_1 < 1$. Another Taylor expansion of ∇f_Y about x gives

$$\nabla f_Y(y; x_j + \xi_j) = \nabla f_Y(y; x) + \nabla^2 f_Y(y; x + \zeta_j)(x_j + \xi_j - x) \qquad (4.4.7)$$

for some vector $\zeta_j = \theta_2(x_j + \xi_j - x)$, $0 < \theta_2 < 1$. Substitution of (4.4.7) into (4.4.6) gives

$$\begin{aligned} E \;=\; & -\int \psi_\nu\{y - T(x)\}\Big\{\langle u - x_j, \nabla f_Y(y; x)\rangle \\ & +\langle u - x_j, \nabla^2 f_Y(y; x_j + \zeta_j)(x_j + \xi_j - x)\rangle\Big\} \; dy. \qquad (4.4.8) \end{aligned}$$

Since

$$\begin{aligned} \nabla\left[\psi_\nu\{y - T(x)\} f_Y(y; x)\right] \;=\; & -\psi_\nu'\{y - T(x)\} f_Y(y; x)\nabla T(x) \\ & +\psi_\nu\{y - T(x)\}\nabla f_Y(y; x), \qquad (4.4.9) \end{aligned}$$

the first term in (4.4.8) may be evaluated as follows

$$\begin{aligned} & \langle x_j - u, \int \psi_\nu\{y - T(x)\}\nabla f_Y(y; x) \; dy\rangle \\ = \; & \langle x_j - u, \nabla \int \psi_\nu\{y - T(x)\} f_Y(y; x) \; dy\rangle \\ & +\langle x_j - u, \nabla T(x)\rangle \int \psi_\nu'\{y - T(x)\} f_Y(y; x) \; dy \\ = \; & \langle x_j - u, \nabla \mathbf{E}\psi_\nu\{Y(x) - T(x)\}\rangle + \langle x_j - u, \nabla T(x)\rangle\gamma_\nu(x) \\ = \; & \langle x_j - u, \nabla T(x)\rangle\gamma_\nu(x), \qquad (4.4.10) \end{aligned}$$

since $\mathbf{E}\psi_\nu\{y - T(x)\} = 0$ and $\gamma_\nu(x) = \int \psi_\nu'\{y - T(x)\} f_Y(y; x) \; dy$ follows from the definition of γ_ν, given in (4.3.12).

To estimate the second term in (4.4.8), note that ψ_ν is bounded and that, by **A3**, $\int \nabla^2 f_Y = O(1)$. Furthermore, $u - x_j = O(n^{-1})$, since $u \in V_{x_j}$, and $x_j - x + \xi_j = O(\frac{k}{n})$, since $n^{-1} = o(\frac{k}{n})$, by **L1** and **L2**. Combining these facts, one obtains

$$|\langle u - x_j, \int \psi_\nu\{y - T(x)\}\nabla^2 f_Y(y; x + \zeta_j)(x_j + \xi_j - x) \; dy\rangle| \le c_2 \frac{k}{n^2} \qquad (4.4.11)$$

for some $c_2 > 0$. Substituting (4.4.10) and (4.4.11) into (4.4.8) now yields the following estimate for E

$$\begin{aligned} |E| \; & \le \; |\langle x_j - u, \nabla T(x)\rangle|\gamma_\nu(x) + c_2 \frac{k}{n^2} \\ & \le \; c_3 n^{-1}\{\gamma_\nu(x) + \frac{k}{n}\} \quad (c_3 > 0) \end{aligned}$$

as required, since ∇T is bounded. □

Lemma 4.4.2 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L3**.

1. *If x, $s \in J^d$, then there exist $\delta_0 > 0$ and $c_0 > 0$ such that for $0 < \delta \leq \delta_0$*

$$\mathbf{E}\psi'_\nu\{Y(s) - T(x)\} \leq f(0; s)g\{T(s) - T(x), \nu\}\{1 + c_0\delta + O(\nu^2)\}$$

where

$$g(a, b) = \frac{\delta + a}{\{(\delta + a)^2 + b^2\}^{1/2}} + \frac{\delta - a}{\{(\delta - a)^2 + b^2\}^{1/2}}.$$

2. *If $x \in J^d$, and s is chosen such that $s = x_j$ for some $j \in L_k(x)$, or if $s = x + hu$ for $u \in J^d$ and $h = (2k+1)/(2n)$, then there exists a constant $c_1 > 0$ such that*

$$\mathbf{E}\psi'_\nu\{Y(s) - T(x)\} \leq c_1 f(0; x)\left\{1 + O(\frac{k}{n}) + O(\nu^2)\right\}.$$

3. *For $\gamma_\nu(x) = \mathbf{E}\psi'_\nu\{Y(x) - T(x)\}$, $x \in J^d$, there exist constants c_1, $c_2 > 0$ and $\delta > 0$ such that for $\nu < \delta$*

$$c_2 f(0; x)\{1 - \frac{\nu^2}{\delta^2}\} \leq \gamma_\nu(x) \leq c_1 f(0; x)\{1 + O(\nu^2)\}.$$

Proof of Lemma 4.4.2

For x, $s \in J^d$, put

$$Y(s) - T(s) = T(s) - T(x) + \epsilon_s = a + \epsilon_s, \tag{4.4.13}$$

where $a = T(s) - T(x)$. Then

$$A \equiv \mathbf{E}\psi'_\nu\{Y(s) - T(x)\} = \int \psi'_\nu(a + \epsilon)f(\epsilon; s)\, d\epsilon. \tag{4.4.14}$$

The density f satisfies a Lipschitz condition at 0, with respect to its first argument ϵ (see **A4**), and we may therefore write

$$f(\epsilon; s) = f(0; s) + d_s(\epsilon).$$

It now follows that there exists $\delta_0 > 0$ such that for $\delta \leq \delta_0$

$$f(\epsilon; s) \leq f(0; s)\{1 + c_1\delta\} \quad \text{for } \epsilon \in [-\delta, \delta],\ c_1 > 0. \tag{4.4.15}$$

For $\delta \leq \delta_0$, A as defined in (4.4.14) now becomes

$$\begin{aligned} A &= \int_{-\delta}^{\delta} \psi'_\nu(a + \epsilon)f(\epsilon; s)\, d\epsilon + \left(\int_{-\infty}^{-\delta} + \int_{\delta}^{\infty}\right) \psi'_\nu(a + \epsilon)f(\epsilon; s)\, d\epsilon \\ &\leq f(0; s)\{1 + c_1\delta\} \int_{-\delta}^{\delta} \psi'_\nu(a + \epsilon)\, d\epsilon \end{aligned}$$

$$+\left(\int_{-\infty}^{-\delta}+\int_{\delta}^{\infty}\right)\psi_\nu'(a+\epsilon)f(\epsilon;s)\,d\epsilon. \tag{4.4.16}$$

Consider

$$\begin{aligned}\int_{-\delta}^{\delta}\psi_\nu'(a+\epsilon)\,d\epsilon &= \nu^2\int_{-\delta}^{\delta}\{(a+\epsilon)^2+\nu^2\}^{-3/2}\,d\epsilon\\ &= \frac{\delta+a}{\{(\delta+a)^2+\nu^2\}^{1/2}}+\frac{\delta-a}{\{(\delta-a)^2+\nu^2\}^{1/2}}.\end{aligned} \tag{4.4.17}$$

To estimate the remaining integrals in (4.4.16), observe that

$$\psi_\nu'(a+\epsilon)\le\nu^2|a+\epsilon|^{-3},$$

and thus, as $\nu\to 0$,

$$\left(\int_{-\infty}^{-\delta}+\int_{\delta}^{\infty}\right)\psi_\nu'(a+\epsilon)f(\epsilon;s)\,d\epsilon\le c_2\frac{\nu^2}{(a+\delta)^3}=O(\nu^2) \tag{4.4.18}$$

for some $c_2>0$. Substitution of (4.4.17) and (4.4.18) into (4.4.16) now yields

$$\begin{aligned}A &\le f(0;s)\{1+c_1\delta\}g(a,\nu)+O(\nu^2)\\ &= f(0;s)g(a,\nu)\{1+c_1\delta+O(\nu^2)\}\end{aligned} \tag{4.4.19}$$

with g as in the statement of the lemma.

Inequality (4.4.19) gives a bound for A for general x, $s\in J^d$. We next consider some specific choices of s for a fixed x.

If $s=x+hu$, for $h=(2k+1)/(2n)$ and $u\in J^d$, or if $s=x_j$, $j\in L_k(x)$, then $a=T(s)-T(x)=O(k/n)$. But k and ν satisfy **L1–L3** and therefore for large n

$$g(a,\nu)=2\{1-\frac{\nu^2}{\delta^2}\}\{1+o(1)\}. \tag{4.4.20}$$

Using this expression for g, one obtains the following bound for A:

$$A\le 2f(0;s)\{1+c_1\delta+O(\nu^2)+o(1)\}\{1-\frac{\nu^2}{\delta^2}\}. \tag{4.4.21}$$

Next observe that

$$f(0;s)=f(0;x)+\langle s-x,\nabla f\{0;x+\theta(s-x)\}\rangle$$

for some θ, $0<\theta<1$, since f is differentiable. Now ∇f is bounded on J^d and hence

$$f(0;s)-f(0;x)=O(s-x)=O(\frac{k}{n}). \tag{4.4.22}$$

Substitution of (4.4.22) into (4.4.21) gives the estimate

$$\begin{aligned} A &\le 2f(0;x)\{1 + c_1\delta + O(\frac{k}{n}) + O(\nu^2) + o(1)\}\{1 - \frac{\nu^2}{\delta^2}\} \\ &\le c_3 f(0;x)\{1 + O(\frac{k}{n}) + O(\nu^2)\} \quad (c_3 > 0). \end{aligned} \tag{4.4.23}$$

This shows part 2.

To complete the proof, recall that

$$\gamma_\nu(x) = \mathbf{E}\psi'_\nu(\epsilon_x) = \mathbf{E}\psi'_\nu\{Y(x) - T(x)\}.$$

Taking $s = x$ in the proofs of parts 1 and 2, the arguments used above now lead to

$$\gamma_\nu(x) \le c_3 f(0;x)\{1 + O(\nu^2)\}.$$

On the other hand, the above arguments show that there exists $c_4 > 0$ such that

$$\gamma_\nu(x) \ge c_4 f(0;x)\{1 - \frac{\nu^2}{\delta^2}\}.$$

□

Proof of Proposition 4.2

Fix $x \in J^d$, and consider for $k \in \mathbb{N}$, $\nu > 0$

$$\begin{aligned} \beta_k^\nu &\equiv \mathbf{E}H_k^\nu\{x, T(x)\} \\ &= \mathbf{E}\sum_{j \in L_k(x)} \alpha_j(x)\psi_\nu\{Y_j - T(x)\} \\ &= h^{-d}\sum_{j \in L_k(x)} \int_{V_{x_j}} \kappa(\frac{u-x}{h})\mathbf{E}\psi_\nu\{Y_j - T(x)\}\, du, \end{aligned} \tag{4.4.24}$$

where

$$\alpha_j(x) = \begin{cases} h^{-d}\int_{V_{x_j}} \kappa(\frac{u-x}{h})\, du & \text{if } V_{x_j - x} \cap [-h,h]^d \supseteq I_n \\ 0 & \text{otherwise} \end{cases}$$

with I_n a closed cube of volume $(2n)^{-d}$, and

$$L_k(x) = \{j \in \mathbb{Z}^d \ : \ \alpha_j(x) \ne 0\}.$$

(See also (4.2.12) and (4.2.13).) For simplicity of notation, put $L = L_k(x)$. Replacing Y_j, $j \in L$, by $Y(u)$ for $u \in V_{x_j}$ leads to the following expression for β_k^ν:

$$\beta_k^\nu = h^{-d}\sum_{j \in L}\int_{V_{x_j}} \kappa(\frac{u-x}{h})\mathbf{E}\psi_\nu\{Y(u) - T(x)\}\, du + \gamma_\nu(x)O(n^{-1}), \tag{4.4.25}$$

since $\mathbf{E}[\psi_\nu\{Y_j - T(x)\} - \psi_\nu\{Y(u) - T(x)\}] \le c_1\gamma_\nu(x)n^{-1}\{1 + O(\frac{k}{n})\}$ for some $c_1 > 0$, by Lemma 4.4.1.

Next fix $j \in L$. Write V_j for V_{x_j} and consider

$$
\begin{aligned}
W_j &\equiv \int_{V_j} \kappa(\frac{u-x}{h})\mathbf{E}\psi_\nu\{Y(u) - T(x)\}\, du \\
&= \int_{V_j} \kappa(\frac{u-x}{h}) \int \psi_\nu\{y - T(x)\} f_Y(y;u)\, dy\, du \\
&= \int \psi_\nu\{y - T(x)\} \int_{V_j} \kappa(\frac{u-x}{h}) f_Y(y;u)\, du\, dy \\
&= \int \psi_\nu\{y - T(x)\} h^d \int_{\overline{V}_j} \kappa(z) f_Y(y; hz + x)\, dz\, dy,
\end{aligned}
$$

where we have used the change of variable $z = (u - x)/h$. Here $\overline{V}_j$ denotes the cube of volume $O(k^{-d})$ which is obtained from V_j by the above change of variable. One now obtains

$$
\begin{aligned}
W_j &= h^d \int_{\overline{V}_j} \kappa(u) \int \psi_\nu\{y - T(x)\} f_Y(y; hu + x)\, dy\, du \\
&= h^d \int_{\overline{V}_j} \kappa(u)\mathbf{E}\psi_\nu\{Y(hu + x) - T(x)\}\, du. \qquad (4.4.27)
\end{aligned}
$$

We now estimate $\mathbf{E}\psi_\nu\{Y(hu + x) - T(x)\}$. For $u \in \overline{V}_j$, put $\xi = hu$. Observe that $\mathbf{E}\psi_\nu\{Y(s) - T(s)\} = 0$ for $s \in J^d$, and therefore

$$
\begin{aligned}
\mathbf{E}\psi_\nu\{Y(\xi + x) - T(x)\} &= \int [\psi_\nu\{y - T(x)\} - \psi_\nu\{y - T(\xi + x)\}]\, f_Y(y; \xi + x)\, dy \\
&= \int \psi_\nu'\{y - T(\xi + x) + \theta\}\{T(\xi + x) - T(x)\} f_Y(y; \xi + x)\, dy \\
&= \tau \int \psi_\nu'\{y - T(\xi + x) + \theta\} f_Y(y; \xi + x)\, dy, \qquad (4.4.28)
\end{aligned}
$$

where $\tau = \{T(\xi + x) - T(x)\}$, and we have used the Taylor expansion of ψ_ν about $\xi + x$ with mean value $\theta = \theta_1\tau$ for some $0 < \theta_1 < 1$. A change from f_Y to f and a Taylor expansion of f about x now leads to

$$
\begin{aligned}
\mathbf{E}\psi_\nu\{Y(\xi + x) - T(x)\} &= \tau \int \psi_\nu'(\epsilon + \theta) f(\epsilon; \xi + x)\, d\epsilon \\
&= \tau \int \psi_\nu'(\epsilon + \theta)\{f(\epsilon; x) + \langle \xi, \nabla f(\epsilon; \xi_2 + x)\rangle\}\, d\epsilon
\end{aligned}
$$

for some $\xi_2 = \theta_2\xi$, $0 < \theta_2 < 1$. For the first term above, Lemma 4.4.2(2) provides the bound

$$
\int \psi_\nu'(\epsilon + \theta) f(\epsilon; x)\, d\epsilon \le c_2 f(0; x)\{1 + O(\frac{k}{n}) + O(\nu^2)\} \quad (c_2 > 0)
$$

for k, ν^{-1} large, since k and ν satisfy **L1–L3**. Observe that ∇f is bounded and $\nabla f \in L^1$ and therefore, as in the proof of Lemma 4.4.2, one may show that

$$
\tau \int \psi_\nu'(\epsilon + \theta)\langle \xi, \nabla f(\epsilon; \xi_2 + x)\rangle\, d\epsilon \le c_3\tau\langle \xi, \nabla f(0; \xi_2 + x)\rangle\{1 + O(\nu^2)\}
$$

for k, ν^{-1} large, and $c_3 > 0$.

Together these two bounds now lead to

$$\begin{aligned}
&\mathbf{E}\psi_\nu\{Y(\xi + x) - T(x)\} \\
&\leq c_4\tau\left[f(0;x)\{1 + O(\frac{k}{n}) + O(\nu^2)\} + \langle\xi, \nabla f(0;\xi_2 + x)\rangle\{1 + O(\nu^2)\}\right] \\
&= c_4\tau\left[f(0;x)\{1 + O(\frac{k}{n}) + O(\nu^2)\} + h\langle u, \nabla f(0;\xi_2 + x)\rangle\{1 + O(\nu^2)\}\right] \quad (4.4.30)
\end{aligned}$$

Substitution of (4.4.27) and (4.4.30) into (4.4.25) now yields

$$\begin{aligned}
h^{-d}\sum_{j\in L} W_j &= \sum_{j\in L}\int_{\overline{V}_j} \kappa(u)\mathbf{E}\psi_\nu\{Y(hu + x) - T(x)\}\, du \\
&= \int_{J^d} \kappa(u)\mathbf{E}\psi_\nu\{Y(hu + x) - T(x)\}\, du \\
&\leq c_4\int_{J^d} \kappa(u)\tau\left[f(0;x)\{1 + O(\frac{k}{n}) + O(\nu^2)\}\right. \\
&\qquad \left. + h\langle u, \nabla f(0;\xi_2 + x)\rangle\{1 + O(\nu^2)\}\right]\, du, \qquad (4.4.31)
\end{aligned}$$

since $h = (2k+1)/(2n)$. Next observe that T has two bounded derivatives (by **A1**) and hence

$$\begin{aligned}
\tau = T(\xi + x) - T(x) &= \langle\xi, \nabla T(x)\rangle + \frac{1}{2}\langle\xi, \nabla^2 T(x + \theta_3\xi)\xi\rangle \\
&= h\langle u, \nabla T(x)\rangle + \frac{1}{2}h^2\langle u, \nabla^2 T(x + \theta_3 hu)u\rangle \qquad (4.4.32)
\end{aligned}$$

for some $0 < \theta_3 < 1$ and $\xi = hu$.

Using the identity (4.4.32) and the fact that $\xi_2 = \theta_2\xi = \theta_2 hu$, (4.4.31) is estimated by

$$\begin{aligned}
h^{-d}\sum_{j\in L} W_j &\leq c_4\int_{J^d} \kappa(u)\{h\langle u, \nabla T(x)\rangle + \frac{1}{2}h^2\langle u, \nabla^2 T(x + \theta_3 hu)u\rangle\} \\
&\quad \times \left[f(0;x)\{1 + O(h) + O(\nu^2)\} + h\langle u, \nabla f(0;\theta_2 hu + x)\rangle\{1 + O(\nu^2)\}\right]\, du \\
&\leq c_5 h^2\left(\left[f(0;x)\{1 + O(h) + O(\nu^2)\}\right]\int_{J^d} \kappa(u)\langle u, \nabla^2 T(x + \theta_3 hu)u\rangle\, du\right. \\
&\quad \left. + \{1 + O(\nu^2)\}\int_{J^d} \kappa(u)\langle u, \nabla T(x)\otimes\nabla f(0;x + \theta_2 hu)u\rangle\, du\right)
\end{aligned}$$

for some $c_5 > 0$, since κ is symmetric. Here $\nabla T(x)\otimes\nabla f(0;x + \theta_2 hu)$ denotes the tensor product of the vectors $\nabla T(x)$ and $\nabla f(0;x + \theta_2 hu)$.

Since $\gamma_\nu(x) \leq c_6 f(0;x)\{1 + O(\nu^2)\}$, by Lemma 4.4.2, since $h = (2k+1)/(2n)$ and $\kappa(u) = 2^{-d}$, we have

$$\beta_k^\nu = h^{-d}\sum_{j\in L} W_j + \gamma_\nu(x)O(n^{-1})$$

$$\leq \; c_7 f(0;x)\{(\frac{k}{n})^2 + n^{-1}\}\{1 + O(\nu^2) + O(\frac{k}{n})\}. \tag{4.4.34}$$

Here

$$c_7 = \max\{c_6, 2^{-d}c_5 C_1, 2^{-d}c_5 C_2\},$$

where

$$C_1 = \sup_{0\leq\theta_3\leq 1} \int_{J^d} \langle u, \nabla^2 T(x + \theta_3 hu)u\rangle \, du$$

$$C_2 = f(0;x)^{-1} \sup_{0\leq\theta_2\leq 1} \int_{J^d} \langle u, \nabla T(x) \otimes \nabla f(0; x + \theta_2 hu)u\rangle \, du.$$

This completes the 'bias' part of the proof.

For $x \in J^d$, $k \in \mathbf{N}$, $\nu > 0$, $j \in L_k(x)$, put

$$Z_j = \psi_\nu\{Y_j - T(x)\}.$$

Consider

$$\begin{aligned} v_k^\nu &= \operatorname{var} H_k^\nu\{x, T(x)\} \\ &= \mathbf{E}\{\sum_{j\in L_k(x)} \alpha_j(x)(Z_j - \mathbf{E}Z_j)\}^2 \\ &= \sum_{j\in L_k(x)} \alpha_j(x)^2\{\mathbf{E}Z_j^2 - (\mathbf{E}Z_j)^2\}. \end{aligned} \tag{4.4.35}$$

Now for $j \in L_k(x)$

$$\begin{aligned} \mathbf{E}Z_j &= \int \psi_\nu\{y - T(x)\} f_Y(y; x_j)\, dy \\ &= \int \psi_\nu\{y - T(x)\} \Big[f_Y(y;x) + \langle x_j - x, \nabla f_Y(y;x)\rangle \\ &\quad + \frac{1}{2}\langle x_j - x, \nabla^2 f_Y\{y; x + \theta_4(x_j - x)\}(x_j - x)\rangle \Big]\, dy \end{aligned} \tag{4.4.36}$$

for some $0 < \theta_4 < 1$. (Recall that f_Y is twice differentiable since f is and ∇f_Y, $\nabla^2 f_Y$ are absolutely integrable by **A3**.)

As in the proof of (4.4.8) in Lemma 4.4.1, we use (4.4.9) to obtain

$$\psi_\nu\{y - T(x)\}\nabla f_Y(y;x) = \nabla\left[\psi_\nu\{y - T(x)\} f_Y(y;x)\right] + \psi_\nu'\{y - T(x)\} f_Y(y;x)\nabla T(x).$$

Substitution into (4.4.36) gives

$$\begin{aligned} \mathbf{E}Z_j &= (1 + \nabla)\int \psi_\nu\{y - T(x)\} f_Y(y;x)\, dy \\ &\quad + \langle x_j - x, \nabla T(x)\rangle \int \psi_\nu'\{y - T(x)\} f_Y(y;x)\, dy \\ &\quad + \frac{1}{2}\int \psi_\nu\{y - T(x)\}\langle x_j - x, \nabla^2 f_Y\{y; x + \theta_4(x_j - x)\}(x_j - x)\rangle\, dy \\ &= (1 + \nabla)\mathbf{E}\psi_\nu(\epsilon_x) + \langle x_j - x, \nabla T(x)\rangle \gamma_\nu(x) + O(|x_j - x|^2) \end{aligned}$$

$$= \gamma_\nu(x)\langle x_j - x, \nabla T(x)\rangle\{1 + O(x_j - x)\}. \tag{4.4.37}$$

This follows since $\mathbf{E}\psi_\nu(\epsilon_x) = 0$, and thus the first term disappears. For the last term we have used the facts that $\psi_\nu(z) \leq 1$ for any $z \in \mathbf{R}$, and $\int \nabla^2 f_Y \; dy = O(1)$.

To estimate the term $\mathbf{E}Z_j^2$ in (4.4.35), note that

$$\begin{aligned} \mathbf{E}Z_j^2 &= \int \psi_\nu\{y - T(x)\}^2 f_Y(y; x_j) \; dy \\ &\leq \int f_Y(y; x_j) \; dy \; = \; 1. \end{aligned} \tag{4.4.38}$$

Substitution of (4.4.37) and (4.4.38) into (4.4.35) leads to

$$\begin{aligned} v_k^\nu &\leq \sum_{j \in L_k(x)} \alpha_j(x)^2 \left[1 + \gamma_\nu(x)^2 \langle x_j - x, \nabla T(x)\rangle^2 \{1 + O(x_j - x)\}\right] \\ &= \sum_{j \in L_k(x)} (2k+1)^{-2d} \left[1 + O\{(\frac{k}{n})^2\}\right] \\ &= (2k+1)^{-d} \left[1 + O\{(\frac{k}{n})^2\}\right] \\ &\leq c_8 k^{-d} \left[1 + O\{(\frac{k}{n})^2\}\right] \quad (c_8 > 0), \end{aligned} \tag{4.4.39}$$

since $x_j - x = O(\frac{k}{n})$, and ∇T is bounded by **A1**.

Equations (4.4.34) and (4.4.39) now yield the estimate:

$$\begin{aligned} \mathbf{E}H_k^\nu\{x, T(x)\}^2 &= (\beta_k^\nu)^2 + v_k^\nu \\ &\leq c_7^2 f(0; x)^2 \left\{(\frac{k}{n})^4 + n^{-2}\right\} \left\{1 + O(\nu^2) + O(\frac{k}{n})\right\} \\ &\quad + c_8 k^{-d} \left[1 + O\{(\frac{k}{n})^2\}\right] \\ &\leq c_9 \left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\} \left\{1 + O(\nu^2) + O(\frac{k}{n})\right\} \end{aligned}$$

as $k \to \infty$, $\nu \to 0$ for some $c_9 > 0$.

This completes the proof of Proposition 4.2. □

4.4.3 Proof of Proposition 4.3

Proposition 4.3 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L4**. *If $x \in J^d$, then*

$$MSE\{\widehat{T}_k^\nu(x)\} \leq c_1 \left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\} \{1 + o(1)\}$$

as $n \to \infty$, for some $c_1 > 0$.

The proof of this proposition requires Lemmas 4.4.3 and 4.4.4, which we state and prove now.

Lemma 4.4.3 *Assume that T and ϵ satisfy* **A1–A3** *and that k satisfies* **L1–L2**. *If $x \in J^d$, then, as $k \to \infty$,*

$$\widehat{T}_k(x) - T(x) = O(\frac{k}{n}) \quad a.s.$$

Proof of Lemma 4.4.3

Fix $x \in J^d$. For n, $k > 0$ there exists $j \in \mathcal{G}_n$ such that $x \in V_{x_j}$. Since $\widehat{T}_k(x) = \widehat{T}_k(x_j)$, by (4.2.17), we put

$$\widehat{T}_k(x) = \operatorname*{med}_{\ell}(Y_{j+\ell})$$

where med_ℓ denotes the median over the set $\{\ell \in \mathcal{G}_n : |\ell^{(i)}| \leq k \;\; \forall i = 1, \ldots, d\}$. Consider

$$\begin{aligned} |\widehat{T}_k(x) - T(x)| &= |\operatorname*{med}_{\ell}\{Y_{j+\ell} - T(x)\}| \\ &= |\operatorname*{med}_{\ell}\{T(x_{j+\ell}) - T(x) + \epsilon_{j+\ell}\}| \\ &= \left|\operatorname*{med}_{\ell}\{\epsilon_{j+\ell} + \langle \xi_\ell, \nabla T(x)\rangle + \frac{1}{2}\langle \xi_\ell, \nabla^2 T(x + \theta\xi_\ell)\xi_\ell\rangle\}\right| \quad (4.4.41) \end{aligned}$$

where $0 < \theta < 1$, $\xi_\ell = x_{j+\ell} - x$, and we have used the Taylor expansion of T about x.

Observe that for any sequence of random numbers a_ℓ, b_ℓ

$$|\operatorname*{med}_{\ell}(a_\ell + b_\ell)| \leq |\operatorname*{med}_{\ell}(a_\ell)| + \max_{\ell}|b_\ell|.$$

Making use of this inequality in (4.4.41) leads to

$$|\widehat{T}_k(x) - T(x)| \leq |\operatorname*{med}_{\ell}(\epsilon_{j+\ell})| + \max_{\ell}|\langle \xi_\ell, \nabla T(x)\rangle + \frac{1}{2}\langle \xi_\ell, \nabla^2 T(x + \theta\xi_\ell)\xi_\ell\rangle|. \quad (4.4.42)$$

Now, $\nabla^2 T$ is bounded and $|\xi_\ell| = |x_{j+\ell} - x| \leq c_1 k/n$ for some $c_1 > 0$, by the definition of med_ℓ. We therefore estimate the second term in (4.4.42) by

$$\begin{aligned} \max_{\ell}|\langle \xi_\ell, \nabla T(x)\rangle + \frac{1}{2}\langle \xi_\ell, \nabla^2 T(x + \theta\xi_\ell)\xi_\ell\rangle| &\leq c_2\frac{k}{n} + c_3(\frac{k}{n})^2 \\ &= O(\frac{k}{n}) \quad (4.4.43) \end{aligned}$$

for c_2, $c_3 > 0$, since $k/n \to 0$ by **L2**.

For the first term in (4.4.42), note that

$$\sum_{\ell \in L_k(x)} \alpha_\ell(x) F_\ell \to F_j$$

as $k \to \infty$, where F_ℓ denotes the distribution function of ϵ_ℓ. An application of the Extended Borel-Cantelli theorem (see p105f of Shorack and Wellner (1986)) together with the proof given on p7 of Pollard (1984) now leads to

$$\operatorname*{med}_{\ell}\{\epsilon_{j+\ell}\} \to \operatorname{med}\epsilon_j \quad \text{a.s., as } k \to \infty.$$

But the probability density function f of ϵ is symmetric by **A2**. From this it follows that

$$\operatorname{med}\epsilon_j = 0. \tag{4.4.44}$$

Substitution of (4.4.43) and (4.4.44) into (4.4.42) now gives

$$\hat{T}_k(x) - T(x) = O(\frac{k}{n}) \quad \text{a.s.}$$

as $k \to \infty$. □

Lemma 4.4.4 *Assume that T and ϵ satisfy* **A1–A4**. *For $x \in J^d$, $k \in \mathbb{N}$, $\nu > 0$, put*

$$\tilde{D}_k^\nu\{x, T(x)\} = \sum_{j \in L_k(x)} \alpha_j(x)\psi_\nu'\{Y_j - T(x) - \eta_j\}$$

where $\eta_j = \theta_j\{T(x) - \hat{T}_k^\nu(x)\}$, $0 < \theta_j < 1$.

If k and ν satisfy **L1–L4**, *then*

$$\tilde{D}_k^\nu\{x, T(x)\} = 2f(0;x)\{1 + o(1)\} \quad \textit{in pr.}$$

as $n \to \infty$.

The proof of this lemma requires some of the arguments leading to Lemma 4.4.2, which precedes the proof of Proposition 4.2 in Subsection 4.4.2.

Proof of Lemma 4.4.4

Fix $x \in J^d$. For $k \in \mathbb{N}$, $\nu > 0$, put $L = L_k(x)$ and

$$\tilde{D}_k^\nu \equiv \sum_{j \in L} \alpha_j(x)\psi_\nu'\{Y_j - T(x) - \eta_j\}$$

where $\eta_j = \theta_j\{T(x) - \hat{T}_k^\nu(x)\}$, $0 < \theta_j < 1$.

We begin with an estimate for $\mathbf{E}\tilde{D}_k^\nu$. We then show that $\lim \mathbf{E}\tilde{D}_k^\nu$ exists, and that for ν and k/n sufficiently small, $\tilde{D}_k^\nu = \mathbf{E}\tilde{D}_k^\nu\{1 + o(1)\}$ in probability.

Fix $j \in L$, put $Z_j = \psi_\nu'\{Y_j - T(x) - \eta_j\}$, and consider

$$\mathbf{E}Z_j = \mathbf{E}\psi_\nu'\{Y_j - T(x) - \eta_j\} = \mathbf{E}\psi_\nu'\{T(x_j) - T(x) - \eta_j + \epsilon_j\}.$$

Observe that $T(x_j) - T(x) = O(\frac{k}{n})$, and

$$\begin{aligned}
|\eta_j| &= \theta_j |T(x) - \widehat{T}_k^\nu(x)| \\
&\leq |T(x) - \widehat{T}_k(x)| + |\widehat{T}_k(x) - \widehat{T}_k^\nu(x)| \\
&= O(\frac{k}{n}) + O\{(\frac{k}{n})^2\} \quad \text{a.s.} \\
&= O(\frac{k}{n}) \quad \text{a.s.} \qquad (4.4.45)
\end{aligned}$$

This last follows from Lemma 4.4.3 together with Proposition 4.1 as $\nu \to 0$, $k \to \infty$. To obtain an estimate for $\mathbf{E}Z_j$, we use arguments similar to those given in the proof of Lemma 4.4.2. However, instead of using $0 < \delta < \delta_0$, here we regard δ as a function of n. Specifically, put $\delta(n) = k^{1+\gamma}/n$, where $\gamma > 0$ is chosen small enough that $\delta(n) \to 0$ as $n \to \infty$. For large n, it follows that $\delta(n) \leq \delta_0$. Furthermore, **L1**–**L4** imply that $\nu = o(\delta)$ and $k/n = o(\delta)$ as $n \to \infty$. In this case, one can show that the factor g used in the statement of Lemma 4.4.2 has the following form:

$$g(a,\nu) = 2\{1 - \frac{1}{n^2}\}\{1 + o(1)\} \quad \text{as } n \to \infty.$$

One now obtains, in analogy with Lemma 4.4.2,

$$\begin{aligned}
\mathbf{E}Z_j &\leq 2f(0;x)\{1 + c_1\delta(n)\}\{1 - \frac{1}{n^2}\}\{1 + o(1)\} \\
&\leq 2f(0;x)\left\{1 + c_1\frac{k^{1+\gamma}}{n}\right\}\{1 + o(1)\} \quad (c_1 > 0). \qquad (4.4.46)
\end{aligned}$$

Similarly one may show that

$$\mathbf{E}Z_j \geq 2f(0;x)\left\{1 - c_2\frac{k^{1+\gamma}}{n}\right\}\{1 + o(1)\} \quad (c_2 > 0). \qquad (4.4.47)$$

Using these last two bounds for $\mathbf{E}Z_j$, it follows that

$$\mathbf{E}\tilde{D}_k^\nu = \sum_{j \in L} \alpha_j(x)\mathbf{E}Z_j \to 2f(0;x) \quad \text{as } n \to \infty. \qquad (4.4.48)$$

It remains to show that

$$\tilde{D}_k^\nu = 2f(0;x)\{1 + o(1)\} \quad \text{in pr.} \qquad (4.4.49)$$

as $k \to \infty$, $\nu \to 0$. To do this, put

$$\begin{aligned}
Z_j^\star &= Z_j - \mathbf{E}Z_j \\
D_k^{\nu\star} &= \sum_{j \in L} \alpha_j(x) Z_j^\star.
\end{aligned}$$

The proof of (4.4.49) proceeds along the following lines. We first show that var $D_k^{\nu\star} \to 0$ as $k \to \infty$, $\nu \to 0$. This will imply that $D_k^{\nu\star} \to 0$ in probability.

Put $b_j = T(x_j) - T(x) - \eta_j$ and consider

$$\begin{aligned}
\mathbf{E}Z_j^2 &= \int \psi_\nu'(b_j+\epsilon)^2 f(\epsilon; x_j)\, d\epsilon \\
&= \nu^4 \int \{(b_j+\epsilon)^2+\nu^2\}^{-3} f(\epsilon; x_j)\, d\epsilon \\
&\leq c_3 f(0;x)\{1+O(\nu^2)\} \quad (c_3>0).
\end{aligned} \tag{4.4.51}$$

Using this estimate, one obtains

$$\begin{aligned}
\text{var } D_k^{\nu\star} &= \mathbf{E}\{D_k^{\nu\star}\}^2 \\
&= \sum_{j\in L} \alpha_j(x)^2\{\mathbf{E}Z_j^2 - (\mathbf{E}Z_j)^2\} \\
&\leq c_4 \sum_{j\in L} \alpha_j(x)^2 \left[f(0;x)\{1+O(\nu^2)\} + f(0;x)^2 \left\{1+O(\frac{k^{1+\gamma}}{n})\right\}\right] \\
&\leq c_4 \max\{f(0;x), f(0;x)^2\} \left\{1+O(\frac{k^{1+\gamma}}{n})\right\} \sum_{j\in L}(2k+1)^{-2d} \\
&\to 0
\end{aligned} \tag{4.4.52}$$

for some $c_4 > 0$, as $k \to \infty$, $\nu \to 0$. In the derivation of (4.4.52) we have used the bound given in (4.4.46), as well as the fact that the weights $\alpha_j = (2k+1)^{-d}$ for $j \in L$. Now, (4.4.52) implies that

$$\lim_{\substack{k\to\infty \\ \nu\to 0}} D_k^{\nu\star} = 0 \quad \text{in pr.}$$

From the last relationship, the definition of $D_k^{\nu\star}$ and (4.4.48), it now follows that

$$\tilde{D}_k^\nu = \mathbf{E}\tilde{D}_k^\nu\{1+o(1)\} \quad \text{in pr.}$$

and therefore

$$\lim_{\substack{k\to\infty \\ \nu\to 0}} \tilde{D}_k^\nu = 2f(0;x) \quad \text{in pr.}$$

This completes the proof of the lemma. □

Proof of Proposition 4.3

Fix $x \in J^d$. For $k \in \mathbf{N}$, $\nu > 0$, put

$$r_k^\nu = \widehat{T}_k^\nu(x) - T(x), \tag{4.4.53}$$

and recall from (4.3.11) that

$$r_k^\nu = H_k^\nu\{x, T(x)\}\left[\tilde{D}_k^\nu\{x, T(x)\}\right]^{-1}. \tag{4.4.54}$$

By Lemma 4.4.4,

$$\tilde{D}_k^\nu\{x,T(x)\} \to 2f(0;x) \text{ in pr.} \tag{4.4.55}$$

as $k \to \infty$, $\nu \to 0$. Since $f(0;x) \neq 0$, it follows that

$$\tilde{D}_k^\nu\{x,T(x)\}^{-1} \to \{2f(0;x)\}^{-1} \text{ in pr.}$$

We may therefore write

$$\begin{aligned} r_k^\nu &= H_k^\nu\{x,T(x)\}\{2f(0;x)\}^{-1}\{1+o(1)\} \text{ in pr.} \\ &= (\mathbf{E}H_k^\nu\{x,T(x)\} + [H_k^\nu\{x,T(x)\} - \mathbf{E}H_k^\nu\{x,T(x)\}]) \\ &\quad \times\{2f(0;x)\}^{-1}\{1+o(1)\} \text{ in pr.} \end{aligned} \tag{4.4.56}$$

Consider the term $H_k^\nu - \mathbf{E}H_k^\nu$. Put $Z_j = \psi_\nu\{Y_j - T(x)\}$ for $j \in L_k(x)$ and recall that

$$H_k^\nu\{x,T(x)\} = \sum_{j\in L_k(x)} \alpha_j(x)\psi_\nu\{Y_j - T(x)\}.$$

The random variables Z_j are uncorrelated and

$$\begin{aligned} \mathbf{E}Z_j^2 &= \int \psi_\nu\{y-T(x)\}^2 f_Y(y;x_j)\,dy \\ &= \int \{y-T(x)\}^2\left[\{y-T(x)\}^2+\nu^2\right]^{-1} f_Y(y;x_j)\,dy \\ &\leq 1. \end{aligned}$$

The strong law of large numbers (see Theorem 5.1.2 of Chung (1974)) now yields

$$\begin{aligned} H_k^\nu\{x,T(x)\} &= \mathbf{E}H_k^\nu\{x,T(x)\}\{1+o(1)\} \text{ a.s.} \\ &= \beta_k^\nu\{1+o(1)\} \text{ a.s., as } k\to\infty,\ \nu\to 0, \end{aligned}$$

and therefore

$$H_k^\nu\{x,T(x)\} - \mathbf{E}H_k^\nu\{x,T(x)\} \to 0 \text{ a.s.}$$

By Proposition 4.2, $\beta_k^\nu \leq c_1\{(\frac{k}{n})^2 + n^{-1}\}\{1+O(\nu^2)+O(\frac{k}{n})\}$, and we may therefore apply Theorem 4.1.4 of Chung (1974) to deduce that

$$\mathbf{E}\,|H_k^\nu\{x,T(x)\} - \mathbf{E}H_k^\nu\{x,T(x)\}|^p \to 0 \text{ for } p=1,2.$$

It now follows that the bias B and the variance V of $\widehat{T}_k^\nu$ are

$$\begin{aligned} B(r_k^\nu) &= \mathbf{E}\{\widehat{T}_k^\nu(x) - T(x)\} \\ &= \mathbf{E}H_k^\nu\{x,T(x)\}\{2f(0;x)\}^{-1}\{1+o(1)\} \\ V(r_k^\nu) &= \text{var}\,\{\widehat{T}_k^\nu(x) - T(x)\} \\ &= \text{var}\,H_k^\nu\{x,T(x)\}\{2f(0;x)\}^{-2}\{1+o(1)\}. \end{aligned} \tag{4.4.59}$$

Applying the results of Proposition 4.2 for expected value and variance of $H_k^\nu\{x, T(x)\}$ yields, as $k \to \infty$, $\nu \to 0$,

$$\begin{aligned} B(r_k^\nu) &\leq c_2\{(\frac{k}{n})^2 + n^{-1}\}\{1 + O(\nu^2) + O(\frac{k}{n}) + o(1)\} \\ V(r_k^\nu) &\leq c_3 k^{-d}\left[1 + O\{(\frac{k}{n})^2\} + o(1)\right] \end{aligned}$$

for some c_2, $c_3 > 0$. The mean square error of $\widehat{T}_k^\nu(x)$ is now estimated by

$$\mathrm{MSE}\{\widehat{T}_k^\nu(x)\} \leq c_4\left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\}\{1 + o(1)\},$$

for some $c_4 > 0$, as $k \to \infty$, $\nu \to 0$, since k, ν satisfy **L1–L4**. This completes the proof of Proposition 4.3. □

4.4.4 Proof of Theorem 4.4

Theorem 4.4 *Assume that T and ϵ satisfy* **A1–A4** *and that k and ν satisfy* **L1–L4**. *If $x \in J^d$, then as $n \to \infty$,*

$$MSE\{\widehat{T}_k(x)\} = O\left\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\right\}. \tag{4.4.61}$$

Furthermore, optimal choices of k and the associated MSE are as follows:

1. *If $1 \leq d \leq 4$, then $k^\star(n) = n^{4/(4+d)}$ minimises the order of MSE and*

$$MSE\{\widehat{T}_{k^\star}(x)\} = O\{n^{-4d/(4+d)}\}. \tag{4.4.62}$$

2. *If $d \geq 4$, then $k^\star(n) = n^{2/d}$ minimises the order of MSE and*

$$MSE\{\widehat{T}_{k^\star}(x)\} = O(n^{-2}). \tag{4.4.63}$$

Proof of Theorem 4.4

For $x \in J^d$, $k \in \mathbb{N}$, $\nu > 0$, write

$$\widehat{T}_k(x) - T(x) = \{\widehat{T}_k(x) - \widehat{T}_k^\nu(x)\} + \{\widehat{T}_k^\nu(x) - T(x)\}$$

and observe that

$$\mathbf{E}\{\widehat{T}_k(x) - T(x)\}^2 \leq c_1\mathbf{E}\{\widehat{T}_k(x) - \widehat{T}_k^\nu(x)\}^2 + c_2\mathbf{E}\{\widehat{T}_k^\nu(x) - T(x)\}^2 \tag{4.4.64}$$

for some c_1 $c_2 > 0$.

Now, since $\nu(n) \leq c_3 k^{-d} n^{-1} \max\{k^2 n^{-1}, 1\}$ for $c_3 > 0$, it follows by Proposition 4.1 that there exists $c_4 > 0$ such that

$$|\widehat{T}_k(x) - \widehat{T}_k^{\nu}(x)| \leq c_4 \left\{ (\frac{k}{n})^2 + n^{-1} \right\}. \tag{4.4.65}$$

The bound given on the right hand side of (4.4.65) is not random, as is clear from the proof of Proposition 4.1, although both $\widehat{T}_k$ and $\widehat{T}_k^{\nu}$ are.

A bound for the second term in (4.4.64) is given by Proposition 4.3 as

$$\mathbf{E}\{\widehat{T}_k^{\nu}(x) - T(x)\}^2 \leq c_5 \left\{ (\frac{k}{n})^4 + n^{-2} + k^{-d} \right\} \{1 + o(1)\}, \tag{4.4.66}$$

for some $c_5 > 0$.

Combining (4.4.65) and (4.4.66), the mean square error of $\widehat{T}_k(x)$ is bounded by

$$\mathrm{MSE}\{\widehat{T}_k(x)\} \leq c_6 \left\{ (\frac{k}{n})^4 + n^{-2} + k^{-d} \right\} \{1 + o(1)\} \tag{4.4.67}$$

$k \to \infty$, $\nu \to 0$, for some $c_6 > 0$.

To obtain optimal rates of convergence for the MSE we choose the parameter k in such a way that it minimises the order of MSE. We distinguish three cases, depending on the dimension d of the observations, which combine to form the two in the statement of the theorem:

1. Take $1 \leq d \leq 3$. Equating the terms $(\frac{k}{n})^4$ and k^{-d} of (4.4.67) leads to

   $$k(n) = n^{4/(4+d)}. \tag{4.4.68}$$

   This choice of k satisfies **L1** and **L2**. Furthermore, $n^{-2} = o\{(\frac{k}{n})^4\}$ implies that the term n^{-1} in the expression for the bias is negligible. For k as in (4.4.68), MSE becomes

   $$\mathrm{MSE}\{\widehat{T}_k(x)\} = O\left\{ (\frac{k}{n})^4 + k^{-d} \right\} = O\{n^{-4d/(4+d)}\}.$$

2. Take $d > 4$. Equating the second term due to bias (n^{-2}) with the variance term k^{-d} yields

   $$k(n) = n^{-2/d}$$

   as the optimal window size. In this case, the first term in the expression of the bias, namely $(\frac{k}{n})^4$, is negligible and

   $$\mathrm{MSE}\{\widehat{T}_k(x)\} = O\{n^{-2} + k^{-d}\} = O(n^{-2}).$$

3. Finally, take $d = 4$. In this case, $k(n) = n^{1/2}$ satisfies all the requirements on k. Furthermore, for this k,

$$k^{-4} = n^{-2} = (\frac{k}{n})^4,$$

that is, all three terms in the mean square error are of the same order, and

$$\mathrm{MSE}\{\widehat{T}_k(x)\} = O\{(\frac{k}{n})^4 + n^{-2} + k^{-d}\} = O(n^{-2}).$$

□

References

Adler, R.J. (1981) *The Geometry of Random Fields.* John Wiley, New York.

Bateman Manuscript Project (1954) *Tables of Integral Transforms, Vol 1.* Ed. A. Erdelyi, McGraw Hill, New York.

Bertero, M. (1986) Regularization methods for linear inverse problems, in *Inverse Problems, CIME.* Ed. G. Talenti, Lecture Notes in Mathematics 1225, Springer-Verlag, Berlin.

Besag, J. (1986) On the statistical analysis of dirty pictures. (With Discussion.) *J. R. Statist. Soc. Ser. B* 48, 259-302

Bloomfield, P. and Steiger, W.L. (1983) *Least Absolute Deviations: Theory, Applications, and Algorithms.* Birkhäuser, Boston.

Bovik, A.C., Huang, T.S., and Munson, D.C. (1983) A generalization of median filtering using linear combinations of order statistics. *IEEE ASSP-31*, 1342-50.

Bovik, A.C., Huang, T.S., and Munson, D.C. (1987) The effect of median filtering on edge estimation and detection. *IEEE PAMI-9*, 181-194.

Brockwell, P.J. and Davis, R.A. (1987) *Time Series: Theory and Methods.* Springer-Verlag, New York.

Chin, R.T. and Yeh, C.-L. (1983) Quantitative evaluation of some edge-preserving noise-smoothing techniques. *Computer Vision, Graphics and Image Processing* 23, 67-91.

Chung, K.L. (1974) *A Course in Probability Theory.* Academic Press, London.

Cramér, H. (1946) *Mathematical Methods of Statistics.* Princeton University Press.

Eubank, R.L. (1988) *Spline Smoothing and Nonparametric Regression.* Marcel Dekker, New York.

Gallagher, N.C. and Wise, G.L. (1981) A theoretical analysis of the properties of median filters. *IEEE ASSP-29*, 1136-41.

Golub, G., Heath, H. and Wahba, G. (1979) Generalized cross-validation as a method for choosing a ridge parameter. *Technometrics* 21, 215-223.

Härdle, W. (1989) *Applied Nonparametric Regression,* Oxford University Press, Oxford.

Härdle, W., and Gasser, T. (1984) Robust non-parametric function fitting. *J. R. Statist. Soc. Ser. B* 46, 242-51.

Härdle, W., Hall, P. and Marron, J.S. (1988) How far are automatically chosen regression smoothing parameters from their optimum? *J. Amer. Statist. Assoc.* 83, 86-95.

Härdle, W., and Marron, J.S. (1985) Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* 15, 1465-1481.

Hall, E.L. (1979) *Computer Image Processing and Recognition.* Academic Press, New York.

Hall, P. (1987a) On the amount of detail that can be recovered from a degraded signal. *Adv. Appl. Prob.* 19, 371-395.

Hall, P. (1987b) On the processing of a motion-blurred image. *SIAM J. Appl. Math.* 47, 441-453.

Hall, P. (1989) On convergence rates in nonparametric problems. *International Statistical Review* 57, 45-58.

Hall, P. (1990) Optimal convergence rates in signal recovery. *Ann. Prob.* 18, 887-900.

Hall, P. and Heyde, C.C. (1980) *Martingale Limit Theory and Its Application.* Academic Press, London.

Hall, P. and Koch, I. (1990) On continuous image models and image analysis in the presence of correlated noise. *Adv. Appl. Prob.* 22, 332-349.

Hall, P. and Koch, I. (1991) On the feasibility of cross-validation in image analysis. *SIAM J. Appl. Math.* To appear.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986) *Robust Statistics: The Approach based on the Influence Function.* John Wiley, New York.

Huang, T.S., Yang, G.J. and Tang, G.Y. (1979) A fast two-dimensional median filtering algorithm. *IEEE ASSP-27*, 13-18.

Huang, T.S. (1981) *Two-Dimensional Signal Processing: Transforms and Median Filters.* Springer-Verlag, New York.

Huber, P.J. (1964) Robust estimation of a location parameter. *Ann. Math. Statist.* 35, 73-101.

Ibragimov, I.A. and Linnik, Yu.V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

Koch, I. and Anderssen, R.S. (1986) Direct surface smoothing for image reconstruction. In *Computational Techniques and Applications: CTAC-85.* Eds. J. Noye and R. May, North-Holland, Amsterdam.

Koch, I. and Tarlowski, C. (1987) A direct surface smoothing procedure for the downward continuation problem. CSIRO DMS Consulting Report ACT 87/8.

Lukas, M.A. (1980) Regularization. In *Application and Numerical Solution of Integral Equations.* Eds. R.S. Anderssen, F.R. de Hoog and M.A. Lukas, Sijthoff and Noordhoff International Publishers.

Phillips, D.L. (1962) A technique for the numerical solution of certain integral equations of the first kind. *J. Assoc. Comput. Mach.* 9, 84-97.

Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer-Verlag, New York.

Priestley, M.B., and Chao, M.T. (1972) Non-parametric function fitting. *J.R. Statist. Soc. B* 34, 385-392.

Reinsch, H. (1967) Smoothing by spline functions. *Numer. Math.* 10, 177-183.

Rosenfeld, A. and Kak, A.C. (1976) *Digital Picture Processing.* Academic Press, New York.

Schoenberg, I.J. (1964) Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.* 52, 947-950.

Serra, J.P. (1982) *Image Analysis and Mathematical Morphology.* Academic Press, New York.

Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* John Wiley, New York.

Silverman, B.W. (1985) Some aspects of the spline smoothing approach to nonparametric curve fitting. (With Discussion.) *J. Roy. Statist. Soc. Ser. B* 47,1-52.

Silverman, B.W. (1986) *Density Estimation in Statistics and Data Analysis.* Chapman and Hall, London.

Stone, C.J. (1984) An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.* 12, 1285-97.

Thompson, A.M., Kay, J.W. and Titterington, D.M. (1990) A cautionary note about crossvalidatory choice. ~~*Preprint.*~~ J. Statist. Comput. Simul. 33, 199-216

Tikhonov, A.N. (1963a) Solutions of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.* 4, 1035-1038.

Tikhonov, A.N. (1963b) Regularization of incorrectly posed problems. *Soviet Math. Dokl.* 4, 1624-1627.

Tikhonov, A.N. and Arsenin, V. (1977) *Solutions of ill-posed problems.* Wiley, New York.

Titterington, D.M. (1985) General structure of regularization procedures in image processing. *Astron. Astrophys.* 144, 381-387.

Tukey, J.W. (1977) *Exploratory Data Analysis.* Addison-Wesley, Reading, Mass.

Twomey, S. (1965) The application of numerical filtering to the solution of integral equations encountered in indirect sensing measurements. *J. Franklin Institute* 279, 95-109.

Vieu, P. (1991) Nonparametric regression: optimal local bandwidth choice. *J. R. Statist. Soc. Ser. B* 53, 453-464.

Vo-Khac Khoan (1972) *Distributions Analyse de Fourier Opérateurs aux Dérivées Partielles, tome I et II,* Librairie Vuibert, Paris.

Wahba, G. (1990) *Spline Models for Observational Data.* CBMS-NSF Regional Conference Series in Applied Mathematics 59, SIAM, Philadelphia.

Yang, G.J. and Huang, T.S. (1981) The effect of median filtering on edge location estimation. *Computer Graphics and Image Processing* 15, 224-245.

Basawa, I.V. and Paraksa Rao, B.L.S. (1980) *Statistical Inference for Stochastic Processes.* Academic Press, London.

Butzer, P.L. and Nessel, R.J. (1971) *Fourier Analysis and Approximation.* Birkhäuser Verlag, Basel.

Doob, J.L. (1953) *Stochastic Processes.* Wiley, New York.

Dudley, R.M. (1989) *Real Analysis and Probability.* Wadsworth & Brooks/Cole, Pacific Grove, California.

Maritz, J.S. (1981) *Distribution-Free Statistical Methods.* Chapman and Hall, London.

Reeves, S.J. and Mersereau, R.M. (1980) Optimal estimation of the regularization parameter and stabilizing functional for regularized image restoration. *Optical Engineering* 29, 446-454.

Rudin, W. (1970) *Real and Complex Analysis.* McGraw-Hill, London.